A Broadly Conserved Deoxycytidine Deaminase Protects Bacteria from Phage Infection

Geoffrey B. Severin[1#@], Brian Y. Hsueh[2#], Clinton A. Elg[3], John A. Dover[1], Christopher R. Rhoades[2], Alex J. Wessel[2], Benjamin J. Ridenhour[4], Eva M. Top[3], Janani Ravi[5], Kristin N. Parent[1], and Christopher M. Waters[2*]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA, 48824

[2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA, 48824

[3]Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, Bioinformatics and Computational Biology Program, University of Idaho, Moscow, Idaho, USA, 83844

[4]Department of Mathematics and Statistical Sciences, University of Idaho, Moscow, Idaho, USA, 83844

[5]Department of Pathobiology and Diagnostic Investigation, Michigan State University, East Lansing, Michigan, USA, 48824

Running Title: A Broadly Conserved Cytidine Deaminase Protects Bacteria from Phage Infection

[#]These authors contributed equally to this work.

@Current address: Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, 48109

[*]**Corresponding Author:**

5180 Biomedical and Physical Sciences

567 Wilson Road

East Lansing, MI 48824

Telephone 517-884-5360

E-mail: watersc3@msu.edu

Keywords: cytidine deaminase, phage, genomic island, toxin-antitoxin, thymineless death, APOBEC

SUMMARY

The El Tor biotype of *Vibrio cholerae* is responsible for perpetuating the longest cholera pandemic in recorded history (1961-current). The genomic islands VSP-1 and -2 are two understudied genetic features that distinguish El Tor from previous pandemics. To understand their utility, we calculated the co-occurrence of VSP genes across bacterial genomes. This analysis predicted the previously uncharacterized *vc0175*, herein renamed **d**eoxycytidylate **d**eaminase ***V**ibrio* (*dcdV*), is in a gene network with *dncV*, a cyclic GMP-AMP synthase involved in phage defense. DcdV consists of two domains, a P-loop kinase and a deoxycytidylate deaminase, that are required for the deamination of dCTP and dCMP, inhibiting phage predation by corrupting cellular nucleotide concentrations. Additionally, DcdV is post-translationally inhibited by a unique noncoding RNA encoded 5' of the *dcdV* locus. DcdV homologs are conserved in bacteria and eukaryotes and our results identify *V. cholerae* DcdV as the founding member of a previously undescribed bacterial phage defense system.

INTRODUCTION

Vibrio cholerae, the etiological agent responsible for the diarrheal disease cholera, is a monotrichous, Gram-negative bacterium found ubiquitously in marine environments [1]. There have been seven recorded pandemics of cholera, beginning in 1817, and the fifth and sixth pandemics were caused by strains of the classical biotype. The seventh pandemic, which began in 1961 and continues today, was initiated and perpetuated by circulating strains of the El Tor biotype. Numerous phenotypic and genetic characteristics are used to distinguish the classical

and El Tor biotypes [2]. It is hypothesized that El Tor's acquisition of two unique genomic islands of unknown origins, named the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and 2) [3], played a pivotal role in El Tor's evolution to pandemicity and the displacement of the classic biotype in modern cholera disease [4].

Combined, VSP-1 and VSP-2 encode ~36 putative open reading frames (ORFs) within ~39 kb (Figs. 1A and S1B) [3, 5–7]. While the majority of the genes in these two islands remain to be studied, it is hypothesized that the biological functions they encode may contribute to environmental persistence [8] and/or the pathogenicity [9] of the El Tor biotype. In support of this idea, VSP-1 encodes a phage defense system encompassing the genes *dncV, capV*, *vc0180* and *vc0181* called the cyclic-oligonucleotide-based antiphage signaling system (CBASS) [10] (Fig. 1A). CBASS limits phage invasion of bacterial populations via a process termed abortive replication whereby upon phage infection cyclic GMP-AMP (cGAMP) synthesis by DncV activates cell lysis by stimulating the phospholipase activity of CapV [10, 11]. During our search for VSP-1 and 2 gene networks, we determined that the gene *vc0175,* renamed herein as **d**eoxy**c**ytidylate **d**eaminase in ***V****ibrio* (*dcdV*), cooccurs in bacterial genomes with *dncV*, suggesting a common function.

We show that *dcdV,* exhibits deoxycytidylate deaminase (DCD) activity, catalyzing the deamination of free deoxycytidine monophosphate (dCMP) substrates to form deoxyuridine monophosphate (dUMP) and is part of the broader Zn-dependent cytosine deaminase (CDA) family of enzymes [12–14]. The activity of DCD enzymes play a vital role in the de novo synthesis of deoxythymidine triphosphate (dTTP) by supplying the dUMP required by thymidylate synthase (TS) to form deoxythymidine monophosphate (dTMP) [12]. CDA enzymes belonging to the APOBEC (Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like) family also play an important role in viral immunity in higher organisms where their catalytic activity is utilized for the deamination of nucleic acids rather than free nucleotide substrates to restrict several types of viruses, such as retroviruses, and retroelements [15–19].

A primary challenge faced by lytic phage is to rapidly replicate many copies of its genome, which requires sufficient nucleotide substrates [20]. During DNA phage infection, total DNA within a bacterium can increase 5-10 fold, illustrating the vast amount of DNA replication that must occur in a short window of time [21, 22]. To accomplish this feat, invading DNA phage often corrupt the delicate balance of enzymatic activity across a host's deoxynucleotide biosynthetic pathways by deploying their own DCD, dUTPase, TS, and ribonucleotide reductase to ensure the appropriate ratio and abundance of deoxyribonucleotides [23–27].

Here we show that DcdV is a dual domain protein consisting of a putative N-terminal P-loop kinase (PLK) and C-terminal DCD domain, and this novel domain architecture is present across the tree of life. Overexpression of DcdV promotes cell filamentation, which has hallmarks of nucleotide starvation resembling thymine-less death (TLD) toxicity [28–31]. Our results demonstrate that ectopic expression of DcdV indeed corrupts the intracellular concentrations of deoxynucleotides and this activity protects bacteria from phage infection. Moreover, we demonstrate that DcdV activity is negatively regulated by a non-coding RNA encoded 5' of the *dcdV* locus [renamed herein as **D**cdV **i**nsensitivity **f**actor in ***V****ibrio* (*difV*)]. Furthermore, *dcdV-difV* systems are widely encoded in bacteria and we show that a subset of them function similarly, establishing cytidine deaminase enzymes as antiphage defense systems in bacteria.

RESULTS

dncV and *dcdV* co-occur in bacterial genomes

To help identify functional interactions within the largely unclassified VSP-1 & 2 genes, VSP island genes were classified into putative "gene networks" or sets of genes that form a functional pathway to accomplish a biological task. Since gene networks often share deep evolutionary history among diverse taxa, we hypothesized that the set of genes in a gene network would co-occur together in the genomes of diverse taxa at a higher frequency than

chance alone would predict. Our software package was named 'Correlogy' inspired by [32] and is described in detail in the materials and methods.

We calculated a Pearson correlation followed by a partial correlation correction between each of the VSP island genes from the same island across the sequenced bacterial domain. This resulting partial correlation correction "w_{ij}" has an output normalized to a range of -1 to 1, with a w_{ij} of -1 revealing homologs of genes *i* and *j* never occur in the same species as opposed to a value of 1 in which homologues of genes *i* and *j* always co-occur in the same species. Previous research using well-classified *Escherichia coli* gene networks showed that partial correlation values w_{ij} > 0.045 were highly correlated with shared biological functions [32]. Using the above-mentioned approach, we calculated a partial correlation value w_{ij} for all genes *i* to *j* in VSP-1 (Supplemental File 1) and VSP-2 (Supplemental File 2). We generated a visualization of the Maximum Relatedness Subnetworks (MRS) showing the single highest w_{ij} value for each VSP gene (Figs. 1B, S1A, 1B).

One of our VSP-1 gene networks centered on *dncV* and identifies the experimentally validated CBASS anti-phage system (Fig. 1B) [10]. Curiously, the putative deoxycytidylate deaminase encoded by *vc0175* ,which we renamed *dcdV*, was also found to co-occur with *dncV* (w_{ij} = 0.147) but not with any of the other CBASS members (w_{ij} < 0.045) (Fig. 1B). Recognizing that co-occurrence of *dncV* with *dcdV* may indicate a shared or common biological function, we sought to understand the biological activity of *dcdV*.

Ectopic expression of *dcdV* induces cell filamentation in the absence of VSP-1

To assess the function of DcdV, we generated growth curves in both wild type (WT) *V. cholerae* and a double VSP island deletion strain (ΔVSP-1/2) over-expressing *dcdV* (pDcdV) or vector control (pVector). DcdV overexpression did not impact WT growth but did reduce growth

yield in the ΔVSP-1/2 background (Fig. 1C). We evaluated the cellular morphology of WT and ΔVSP-1/2 strains after overexpression of DcdV and observed expression from pDcdV in the ΔVSP-1/2 background yielded filamentous cell morphologies, suggesting these cells have a defect in cell division that manifests in a reduced growth yield (Fig. 1D). We performed the same image analysis in single island mutants (ΔVSP-1 and ΔVSP-2) and found that the mean cell length increased significantly upon DcdV overexpression only in cells lacking VSP-1 (Fig. 1E). Likewise, overexpression of pDcdV in a laboratory strain of *E. coli* also induced cell filamentation that was inhibited by provision of a single copy cosmid containing VSP-1 (pCCD7) but not the vector cosmid control (pLAFR) (Figs. S2A and S2B). The spiral nature of *V. cholerae* filaments (Fig. 1D) is due to the natural curvature of *V. cholerae* mediated by *crvA* [33, 34]. Taken together, these results indicated that DcdV overexpression severely impacts cell physiology in the absence of VSP-1.

DifV is encoded immediately 5' of the *dcdV* locus in VSP-1

To identify the negative regulator of DcdV activity encoded in VSP-1, we generated partial VSP-1 island deletions and quantified cell filamentation following DcdV expression. Three sections of VSP-1; *dcdV-vc0176*, *vspR-vc0181,* and *vc0182-vc0185,* were individually deleted. Of the three partial VSP-1 deletion strains, expression of pDcdV only induced filamentation in the Δ*dcdV-vc0176* mutant (Fig. 2A). Individual gene deletion mutants of *dcdV* and *vc0176* maintained WT cell morphology following expression of DcdV (Fig. 2B), suggesting the 504 nt intergenic region between *dcdV* and *vc0176* is the source of DcdV inhibition. We identified a 222 nucleotide (nt) open reading frame we named ig^{222} encoded in the same orientation immediately 5' of *dcdV* as a possible candidate for the DcdV regulation (Fig. 1A). Overexpression of DcdV in the Δig^{222} mutant led to cell filamentation (Fig. 2B). Furthermore, complementation of ig^{222} co-expressed from a second plasmid in the Δig^{222} strain prevented DcdV induced filamentation (Fig. 2B). We conclude that ig^{222} contains the necessary genetic

components for inhibiting DcdV activity and refer to this negative regulator as DifV for **D**cdV **i**nsensitivity **f**actor in ***V****ibrios*.

As *dcdV* and *dncV* cooccur in a gene network (Fig. 1B), we hypothesized that the role of DncV was to inactivate DifV, leading to the liberation of DcdV activity. However, co-expression of both DncV and DcdV did not liberate DcdV activity as these cells were not filamentous (Fig. S3). The Δ*ig*^{222} mutant is not filamentous in the absence of pDcdV expression which is likely due to a polar effect originating from the deletion of *ig*^{222}. Indeed, *dcdV* expression was reduced at all growth phases in the Δ*ig*^{222} mutant (Fig. S4).

DifV is an sRNA that post-translationally regulates the activity of DcdV

The fact that DifV inhibits DcdV expressed from a plasmid with exogenous transcription and translation start sites suggests DifV regulates DcdV at a post-translational level. To test this hypothesis, we expressed a *dcdV* C-terminal 6x histidine tagged construct (DcdV^{6xHIS}) in WT and Δ*ig*^{222} *V. cholerae* and probed for the cellular abundance of DcdV^{6xHIS} using Western blot (Fig. 2C). When this tagged DcdV is expressed, Δ*ig*^{222} manifest a filamentation phenotype while the WT strain does not, indicating the 6x histidine tag does not change the activity of DcdV nor does it inhibit the ability of DifV to regulate DcdV (Fig. S5). Despite the lack of filamentation in the WT strain, the cellular abundance of DcdV^{6xHIS} was slightly greater than Δ*ig*^{222} with an average signal intensity ratio WT:Δ*ig*^{222}) of 1.5 ± 0.3 across three biological replicates, although this difference was not statistically significant. This result connotes that DifV limits DcdV activity after it has been translated and not by reducing the abundance of DcdV.

Given that DifV regulates the activity of DcdV at the post-translational level, we wondered if DifV was a small peptide or an untranslated small regulatory RNA (sRNA). Mutation of the *ig*^{222} rare CTG start codon to a TAG stop codon (222 nt^{STOP}) did not abrogate the ability of this construct to inhibit DcdV activity in trans when co-expressed in the Δ*ig*^{222} strain (Fig. 2D). We then examined a 174 nt ORF completely encoded within *ig*^{222} (174 nt) and found it was also

sufficient to prevent DcdV induced filamentation (Fig. 2D). Additionally, expression of this 174 nt ORF from constructs either lacking a ribosome binding site (174 nt^{-RBS}) or where the native ATG start codon was mutated to a TAA stop codon (174 nt^{STOP}) each retained the ability to inhibit DcdV activity (Fig. 2D). We also identified an ATG start codon on the interior of the 174 nt ORF corresponding to an alternative reading frame and mutation of this interior start codon to a TAA stop codon (174 $nt^{InteriorSTOP}$) also failed to abrogate DifV inhibition of DcdV activity (Fig. 2D). Together, these results suggest that translation of a gene product originating from within ig^{222} is not necessary for DifV activity.

To identify the minimum functional size of *difV* we further truncated this 174 nt segment from both the 5' and 3' ends and found that removal of either 18 bp from the 5' end or 4 bp from the 3' end was sufficient to abolish DifV activity (Fig. 2D). Additionally, expression in trans of npcR_3991 [35], a 104 nt non-coding RNA of unknown function contained within ig^{222}, was also unable to inhibit DcdV filamentation (Fig. 2D). Collectively, these results suggest that DifV is a regulatory RNA that is between 152 to 174 nt long encoded 5' of the *dcdV* locus, and we will therefore refer to the 174 nt locus as *difV* for the remainder of these experiments.

DifV and DcdV constitute a two gene operon that resembles a Toxin-Antitoxin System

The genomic orientation and proximity of *difV* to *dcdV* suggest they may constitute an operon and two previous genome-wide transcriptional start site (TSS) analyses previously identified a common putative TSS 5' of *difV* [36, 37]. To test if *difV* and *dcdV* are indeed expressed as an operon, we performed diagnostic PCR with primers located within *difV* and *dcdV* on cDNA generated from both WT and Δig^{222} RNA (Fig. 3A). As expected, *dcdV* was detected in the cDNA generated from each strain while *difV* was only amplified using the WT cDNA template (Fig. 3B). The presence of an 839 nt PCR product amplified using primers spanning *difV* to *dcdV* from the WT cDNA template, that was not present with Δig^{222} cDNA, confirmed that both genes are present on a shared transcript (Fig. 3B). Additionally, we

quantified the relative abundance of *difV* and *dcdV* RNA using qRT-PCR and found the *difV* locus was approximately 40-, 20-, and 60-fold more abundant than *dcdV* at early exponential, late exponential, and stationary phases, respectively (Fig. 3C). While having several unique features, the co-transcription of *difV* and *dcdV* and the post-translational regulation of DcdV activity by the abundant sRNA DifV resembles Type III Toxin-Antitoxin systems [38].

DcdV induced filamentation requires conserved features of both the PLK and the CDA domains

DcdV is a 532 amino acid polypeptide composed of two putative domains: an unannotated N-terminal domain and a DCD-like C-terminus (Figs. 4A, 4B). Analysis of the N-terminal domain using Pfam did not reveal any conserved domains. However, Phyre2 [39] and PSI-BLAST searches combined with InterProScan [40, 41] analyses revealed that the N-terminus contained features of the P-loop containing nucleoside triphosphate hydrolase (IPR ID: IPR027417) *aka* P-loop kinase (PLK) enzyme family (Figs. 4A,B and S6). PLKs catalyze the reversible phosphotransfer of the γ-phosphate from a nucleotide triphosphate donor to a diverse group of substrates, depending on the enzyme class, including deoxynucleotide monophosphates. Three structural features commonly found in these enzymes include a P-loop/Walker A motif {GxxxxGK[ST]}, a two helical LID module that stabilizes the donor nucleotide triphosphates, and a Walker B motif {hhhh[D/E], where “h” represents a hydrophobic residue} that is partly involved in coordinating Mg^{2+} [42, 43]. Interrogation of the Phyre2 DcdV model (Fig. 4A), InterProScan predictions, and PSI-BLAST primary sequence alignments (Fig. S6) revealed these three features are likely present in the N-terminal domain, suggesting the N-terminus of DcdV is a PLK domain involved in binding nucleotide substrates and performing a phosphotransfer reaction. The C-terminal DCD domain contains a highly conserved zinc-dependent CDA active site motif ([HAE]X_{28}[PCXXC]) (Figs. 4A,B and S6). The constellation of residues that make up the Zn^{2+} binding pocket is composed of three critical amino acids in

DcdV; H382, C411, and C414. Zn^{2+} is required for the catalytic deprotonation of water by a conserved glutamate residue (E384 in DcdV) for the hydrolytic deamination of a cytosine base to uridine.

Hypothesizing that one of the two domains present in DcdV is responsible for cell filamentation in the absence of *difV,* we made site-specific mutations in the conserved residues predicted to be essential for activity in both the PLK and DCD domains. Two variant constructs were generated in the PLK domain targeting the Walker A motif ($DcdV^{S52K}$) and the Walker B motif ($DcdV^{D162A + Q163A}$) (Fig. 4B). Two variants were constructed in the DCD active site; a double substitution of both C411A and C414A ($DcdV^{C411A + C414A}$) to abrogate Zn^{2+} binding and an E384A substitution ($DcdV^{E384A}$) to inhibit the deprotonation of water required for the hydrolytic deamination of cytosine (Fig. 4B). Unlike WT DcdV ($DcdV^{WT}$), all four of the variants failed to induce filamentation when ectopically expressed in *E. coli* (Fig. 4C). The cellular abundance of these variants is comparable to WT DcdV (Fig. S7). This result shows both DcdV domains are necessary for induction of filamentation.

We performed a genetic screen to identify DcdV variants whose activity was no longer inhibited by DifV by expressing a random library of *dcdV* mutants in a Δ*dcdV* mutant strain where *difV* remains intact. Ectopic expression of WT *dcdV* in a Δ*dcdV* mutant does not induce filamentation (Fig. 4D) or produce small, wrinkled colonies on solid agar due to the genomic copy of *difV*. However, *dcdV* mutants that are insensitive to *difV* exhibit a small colony phenotype. Screening ~ 15,000 potential mutants, we identified five unique *dcdV* mutations that encoded single amino acid substitutions (E123K, A126T, K201R, K511E, and Q514R) located in both the PLD and DCD domains that rendered DcdV insensitive to DifV inhibition (Figs. 4B, 4D). Based on the Phyre2 DcdV structural model, all five residues are located on the exterior of the protein (Fig. 4A) suggesting they may be involved in mediating molecular interactions between DifV and DcdV. The only mutation found within a conserved domain feature was the seemingly

innocuous K201R substitution, which is modeled to lie between the two helices of the PLK LID module (Fig. 4A).

DcdV induced filamentation is due to impaired genome replication

Filamentation is a phenotype often associated with thymineless death (TLD) [28] due to nucleotide starvation. A hallmark of TLD is an increased genomic origin to terminus (*ori*/*ter*) ratio resulting from repeated attempts to initiate replication from the origin that ultimately fail to reach the terminus due to a lack of dTTP substrate [44]. Hypothesizing that DcdV induced filamentation may be a consequence of replication inefficiency, analogous to TLD, we measured the *ori/ter* ratio of *V. cholerae* chromosome 1 from WT and Δ*ig*^{222} *V. cholerae* grown to stationary phase overexpressing WT DcdV or a vector control. There was no significant difference in the *ori/ter* ratios following ectopic expression of WT DcdV in WT *V. cholerae* (Fig. 4E), consistent with the observation that these strains do not filament (Figs. 1D and 1E). However, ectopic expression of WT DcdV in the Δ*ig*^{222} mutant, which lacks *difV*, resulted in an *ori*/*ter* ratio ~ 3 times greater than the vector control (Fig. 4F), consistent with cell filamentation (Fig. 2B). We also measured the *ori*/*ter* ratio of the Δ*ig*^{222} mutant expressing *dcdV* with mutations in the PLK or DCD domain. In agreement with an inability to induce filamentation (Fig. 4C), the *ori*/*ter* ratio of these variants was not significantly different from the empty vector control (Fig. 4F). Therefore, DcdV corruption of DNA replication is dependent upon both the PLK and DCD domains.

DcdV catalyzes the deamination of both dCMP and dCTP

Based on the TLD-like genome instability driven by DcdV, we hypothesized this enzyme deaminates free nucleic acid substrates. Though we determined DcdV and DcdV variants were readily retained in *E. coli* lysates (Fig. S7), numerous attempts to purify active DcdV were unsuccessful. This suggested that an unknown cofactor or cellular condition may contribute to

the activity of DcdV that was missing in our purification conditions. Soluble lysates from *E. coli* ectopically expressing DcdV or the DCD active site variant DcdV^{E384A} were supplemented with amine containing nucleotides and monitored for the evolution of NH_4^+, a product of nucleotide deamination. Lysates containing DcdV evolved significantly more ammonium when incubated with dCMP and dCTP, which was not detected in lysates containing the DCD active site variant DcdV^{E384A} (Fig. 5A).

DCD enzymes are unique among the CDAs for their allosteric regulation by both dCTP and dTTP which activate and repress the catalytic deamination of dCMP, respectively, through a G[Y/W]NG allosteric site motif [45, 46]. Such allosteric regulation ensures that nucleotide homeostasis is maintained even if DCD enzymes are present. The allosteric site found in DcdV is composed of a divergent GCND motif suggesting allosteric regulation by dNTPs may not be preserved. In support of this, the deamination of both dCMP and dCTP by soluble lysates containing DcdV were not inhibited by the addition of equimolar dTTP (Fig. S8).

To further understand the catalytic activity of DcdV we spiked 1 µM dCTP into soluble lysates collected from *E. coli* ectopically expressing either WT DcdV or a vector control and quantified the concentrations of dUTP and dUMP over 30 minutes using UPLC-MS/MS. Following addition of 1 µM dCTP the concentrations of both dUTP (Fig. 5B) and dUMP (Fig. 5C) increased in lysates containing DcdV within the first minute while those found in vector control lysates did not dramatically change over the course of the entire experiment. The concentration of dUTP in DcdV containing lysates peaked after five minutes and slowly receded over time (Fig. 5B) while the concentration of dUMP in these lysates continued to increase to a final concentration of ~ 1 µM after 30 minutes (Fig. 5C). Importantly, the equimolar stoichiometry of the1 µM dCTP substrate spike and the 1 µM dUMP detected at the conclusion of the experiment demonstrates that DcdV does not modify nucleotides in a unique manner which would alter their mass. Together these experiments indicate that DcdV deaminates both dCTP

and dCMP substrates and DcdV containing lysates ultimately funnel dCTP to dUMP, indicating DcdV is likely to have profound effects on intracellular nucleotide metabolism.

DcdV decreases intracellular dCTP, dCMP, and dUTP in *E. coli*

Our genetic and in vitro evidence suggested that DcdV catalyzes the deamination of both dCMP and dCTP to the detriment of DNA replication. To quantify the impact of DcdV activity on the intracellular concentrations of deoxyribonucleotide species, we overproduced DcdV, DcdV^{S52K}, DcdV^{E384A}, and an empty vector control in *E. coli* and measured the abundance of these molecules by UPLC-MS/MS. While all strains contained similar levels of dATP, dGTP, dTTP, and dUMP, the intracellular abundance of dCTP and dCMP were significantly reduced in *E. coli* expressing WT DcdV (Figs. 5D, S9). No dUTP was found following expression of WT DcdV while trace amounts of dUTP were detected in the vector and the two DcdV variant strains (Figs. 5D, S9). Unlike the results observed with the in vitro DcdV lysates (Fig. 5C), no increase in intracellular dUMP concentrations were observed when DcdV was expressed. We speculate the difference between dUMP detected in lysate versus in vivo extractions are due to compensatory metabolic pathways active in live cells which are lost in the lysates. Similar results were obtained when a DcdV homolog derived from enterotoxigenic *E. coli* (DcdV^{ETEC}), discussed later in this study, was overexpressed in the same heterologous *E. coli* host (Figs. 5D, S9). Importantly, inactivating amino acid substitutions in conserved features of the PLK (DcdV^{S52K}) or DCD (DcdV^{E384A}) domains blocked DcdV activity, indicating both domains are necessary for the DcdV dependent depletion of intracellular dC pools (Figs. 5D, S9).

Conservation and evolution of DcdV

To identify if DcdV is widely conserved, we used six DcdV homologs as starting points from *V. cholerae*, *Vibrio parahaemolyticus*, *E. coli*, *Proteus mirabilis*, *Aeromonas veronii*, and *Enterobacter cloacae* to perform homology searches across the tree of life (see Methods). We

used a combination of protein domain and orthology databases, homology searches, and multiple sequence alignment for detecting domains, signal peptides, and transmembrane regions to reconstruct the domain architectures of the query DcdV proteins (Fig. S6). In agreement with the Phyre2 model of *V. cholerae* DcdV (Fig. 4A), we identified two distinct domains in all six DcdV homologs, the N-terminal PLK domain and the C-terminal DCD domain (Fig. S6).

We identified numerous homologs containing the core PLK+DCD architecture as well as other variations, which included multiple PLK domain fusions in proteobacteria (*e.g., Klebsiella*, *Vibrio*) and a nucleic acid binding domain (*e.g., Mannheimia*, *Bibersteinia*) (Table S4). Homologs of DcdV were identified in multiple bacterial phyla including Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes (Figs. 6A, a few dominant clusters of homologs are labeled). Interestingly, we found DcdV-like proteins in Archaea (*e.g.,* Thaumarchaeota) and Eukaryota (*e.g.,* Ascomycota) (Figs. 6A, Table S4). While the percentage similarity is ~50% and <30% for archaeal and eukaryotic homologs, respectively, we note these contain comparable domain architectures to the query proteins (Table S4).

Identification and evaluation of Gram-negative DcdV-DifV system homologs

To evaluate the conservation of enzymatic activity we selected three of the core DcdV homologs used in the initial homolog search; *V. parahaemolyticus* O1:Kuk FDA_R31*, P. mirabilis* AR379*,* and *E. coli* H10407 ETEC (Figs. 6A and S10). Expression of all three DcdV homologs in *E. coli* resulted in filamentous cells analogous to *V. cholerae* DcdV (Fig. 6B). These *dcdV* homologs are encoded 3’ of a small ORF, annotated as a hypothetical protein, in an orientation, size, and proximity consistent with *V. cholerae difV*. While there was no strong amino acid or nucleotide sequence similarity among the small ORFs 5’ of the *dcdV* homologs (Figs. S11 and S12) we hypothesized these could encode cognate *difV* negative regulators. Consistent with the inhibition of DcdV activity by DifV from *V. cholerae*, co-expression of the

corresponding DifV with its DcdV partner suppressed the cell filamentation phenotype (Fig. 6B). Additionally, overexpression of $DcdV^{ETEC}$ in a heterologous *E. coli* host also decreased the intracellular concentrations of dCMP, dCTP, and dUTP (Fig. 5D), indicating the catalytic activity of these DcdV homologs are analogous to *V. cholerae* DcdV.

To determine if DifV and the three ORFs encoded upstream of *dcdV* homologs could provide cross-species inhibition of DcdV, we challenged each of the four *dcdVs* with each of the four *difVs* in *E. coli* and looked for DcdV dependent filamentation. Cross-species inhibition of DcdV induced filamentation was observed between *V. parahaemolyticus* and *V. cholerae* when each species' *difV* was expressed in trans (Fig. 6C). However, *difV* from *P. mirabilis* and *E. coli* ETEC were only able to inhibit the activity of their own cognate DcdV (Fig. 6C). These data suggest that while the general mechanism of DifV inhibition of DcdV activity is conserved the specific molecular interactions that mediate this process are not.

Ectopic expression of DcdV reduces phage titers and slows predation

We initiated studies of *dcdV* based on our discovery that this gene co-occurs in bacterial genomes with *dncV*, a critical member of the CBASS antiphage abortive infection system [10, 47]. Additionally, cytidine deaminases are conserved anti-viral defense mechanisms in eukaryotes [15, 17, 48]. These connections led us to hypothesize that DcdV can also provide phage defense by manipulating cellular nucleotide concentrations. To test this hypothesis, we challenged *V. cholerae* WT and Δ*dcdV* with two *V. cholerae* lytic phage with dsDNA genomes, ICP1 and ICP3 [49, 50]. However, we observed no differences in the ability of these phages to kill *V. cholerae* in these conditions (Figs S13A and S13B).

Because ICP1 and ICP3 have coevolved with El Tor *V. cholerae,* it is likely that these phages have evolved mechanisms to counteract *dcdV*. Such resistance to other *V. cholerae* phage defense mechanisms by ICP-1 has been previously demonstrated [51–53]. Therefore, we selected the heterologous host *Shigella flexneri,* a Gram-negative human pathogen, and its

bacteriophage Sf6 , a dsDNA phage from the *Podoviridae* family [54, 55], as a naïve host-phage pair to test the antiphage activity of DcdV and its homologs. Ectopic expression of *dcdV* or its homologs did not impact the growth of *S. flexneri* before the onset of phage killing at ~110 minutes (Figs. 7A-D). *S. flexneri* strains ectopically expressing *dcdV* or its homologs delayed the onset of population collapse caused by Sf6 predation, although the impact of the *V. cholerae* DcdV was more modest than the other three homologs (Figs. 7A-D). Additionally, induction of all four DcdV homologs significantly reduced Sf6 progeny following infection compared to the control strains lacking induction of DcdV (Fig. 7E). Together, these data indicate that DcdV enzymes confer defense against phage infection by delaying population collapse and reducing the proliferation of viable phage progeny.

DISCUSSION

Uncovering the contributions to bacterial fitness of the ~36 genes encoded within the El Tor *V. cholerae* VSP-1 and 2 genomic islands may help elucidate the longevity and persistence of the seventh cholera pandemic. Our bioinformatic approach using Correlogy accurately identified a gene network composed of the VSP-1 antiphage CBASS system (*capV-dncV-vc0180-vc0181*). Interestingly, this also revealed *dncV* is frequently found in genomes with the previously uncharacterized gene *dcdV*. The only function previously ascribed to *dcdV* was an undefined involvement in quorum sensing controlled *V. cholerae* aggregate formation [56].

We showed that DcdV contains a functional DCD domain that catalyzes the deamination of deoxycytidine nucleotides and a putative PLK-like domain of unknown function. We further demonstrate that homologs of this protein are present across the tree of life. Collectively, both domains are required for DcdV to disrupt deoxynucleotide pool homeostasis, which impairs DNA replication and manifests in a filamentous cell morphology. DcdV activity is post-translationally regulated by DifV, a sRNA encoded immediately 5’ of the *dcdV* locus in VSP-1, though the details of this inhibition remain to be fully elucidated. Finally, we demonstrate that

412 DcdV and a set of homologs from other Gram-negative bacteria confer phage resistant
413 properties when expressed in a heterologous host.

414 Cell filamentation is a hallmark of TLD, observed in bacteria and eukaryotes, which
415 arises from a sudden loss of thymine during robust cellular growth [31]. Interestingly, this
416 phenomenon is not limited to dTTP as dGTP starvation elicits a similar response in *E. coli* and is
417 also hypothesized to occur when other deoxynucleotide substrates become disproportionately
418 scarce [29]. In the case of DcdV, it is conceivable the observed filamentation phenotype is a
419 consequence of a TLD-like reduction in dCTP pools that can be termed '*cytosineless death*'.
420 However, while DcdV activity also reduces the intracellular dC pool, it did not significantly
421 increase the intracellular concentrations of dTTP or dUMP in vivo, suggesting a cellular
422 compensatory pathway to combat DcdV activity is at work in intact cells. We speculate that the
423 DCD and PLK domains of DcdV are responsible for this conversion of dC nucleotides to dUMP
424 observed in the bacterial lysates, but we cannot rule out the contribution of other unknown
425 cellular factors. The deamination of dCTP is canonically performed by non-zinc dependent
426 enzymes [57] making the dual substrate repertoire of dCMP and dCTP in DcdV a rare trait.

427 The delicate balance of enzymatic activity across the pyrimidine biosynthesis pathway
428 can be corrupted by viruses that deploy their own DCD, dUTPase, and TS enzymes to hijack
429 host nucleotide biosynthesis to ensure the appropriate ratio and quantities of
430 deoxyribonucleotide precursors for replicating their own genomes [23, 24, 26, 27]. For example,
431 biDCD from chlorovirus PBCV-1, the only DCD previously reported to deaminate both dCMP
432 and dCTP substrates, rapidly catalyzes the conversion of host dC nucleic acids into dTTP thus
433 aiding replication of the A+T rich viral genome [23]. biDCD is allosterically regulated by dCTP
434 and dTTP to activate and inactivate the deaminase, respectively. This regulation provides a
435 means to fine-tune the pool of available dNTPs by preventing the enzyme from deaminating all
436 available dC substrates. Interestingly, DcdV does not appear to have maintained the allosteric
437 nucleotide binding site nor does excess dTTP added to cell lysates alter the catalytic activity of

DcdV towards dCMP or dCTP (Fig. S8), and we propose these differences in enzyme activity are consistent with the function of DcdV as a phage defense mechanism that inhibits phage replication by corrupting cellular nucleotide pools (graphical abstract). Altering pools of available nucleotides has been shown to fend off biological attacks. For example, prokaryotic viperins protect against T7 phage infection by producing modified ribonucleotides that ultimately inhibit phage polymerase-dependent transcription [58]. The SAMHD1 phosphohydrolase enzyme in eukaryotes also inhibits viral infections by depleting cellular nucleotide pools, although its structure and activity are different than DcdV [59–61].

In lieu of a conserved deoxynucleotide allosteric site, DcdV is regulated post-translationally by the DifV untranslated RNA, which is unique among the CDA-family. The spacing, orientation, and relationship of *difV* and *dcdV* may have adapted to perform functions in a manner analogous to Type 2 and Type 3 Toxin-Antitoxin (TA) systems found across the bacterial phyla of which some are involved in antiphage defense and bacterial stress response [62]. While the RNA antitoxin of Type 3 TA systems encode nucleotide repeats [62] no repeat sequences are obvious in DifV indicating that DcdV/DifV may constitute a new TA class. We hypothesize that DcdV is activated upon phage infection by disruption of DifV inhibition, and we are currently preforming experiments to test this hypothesis (graphical abstract). Our systemic search for DcdV homologs containing at least a single PLK and DCD domain revealed hundreds of examples in a variety of bacteria beyond the Proteobacteria phylum including Bacteroidetes and Actinobacteria and a few homologs in archaea and eukaryota.

Phage defense mechanisms are often found clustered together in mobile genetic elements called defense islands [63, 64] and we speculate that the co-occurrence of DcdV and DncV (along with the rest of the CBASS system) in bacterial genomes is a result of their shared anti-phage activity. Our results indicate that DcdV reduces the available dC pool, and we hypothesize that this activity delays phage genome replication potentially decreasing phage burst size. Although the *S. flexneri* host population expressing DcdV eventually collapses, we

speculate that the delay in phage replication could provide an opportunity to prompt other phage defense systems, such as CBASS or a restriction modifications system to further target invading phages [65, 66].

Our study reveals that bacteria, like eukaryotes, also use CDA enzymes to protect against biological invasion although through different mechanisms. The eukaryotic APOBEC proteins deaminate ssRNA, leading to increased mutation and decreased genome stability of RNA viruses, whereas the substrates of DcdV are free deoxynucleotides. Further studies are required to determine if these two biological defense systems evolved from a common CDA ancestor.

ACKNOWLEDGEMENTS

We thank Shannon Manning (STEC Center, Michigan State University), Jessica Jones (U.S FDA), and Allison Brown (U.S. CDC), for providing us with *E. coli* ETEC, *P. mirabilis,* and *V. parahaemolyticus* strains, respectively. We thank Wei Leung Ng (Tuft University) for providing us *V. cholerae* ICP1 and ICP3 phages. We thank Kefei Yu and Dohun Pyeon for valuable suggestions and Dan Jones and Lijun Chen from the MSU RTSF mass spectrometry facility core for their technical support. This work was supported by National Institutes of Health (NIH) grants GM109259, GM110444, GM139537, AI143098, and AI158433 and National Science Foundation (NSF) grant DBI-0939454to C.M.W, the NIH grant GM110185 and the NSF CAREER Award 1750125 to K.N.P., and the NSF Graduate Research Fellowship Grant No. 1842399 to C.A.E. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

MATERIALS AND METHODS

The strains, plasmids, and primers used in this study are listed in Supplementary Table S1, S2, and S3, respectively. Unless otherwise stated, cultures were grown in Luria-Bertani (LB) at 35°C and supplemented with the following as needed: ampicillin (100 µg/mL), kanamycin (100 µg/mL), tetracycline (10 µg/mL), and isopropyl-β-D-thiogalactoside (IPTG) (100 µg/mL). *E. coli* BW29427, a diaminopimelic acid (DAP) auxotroph, was additionally supplemented with 300 µg/mL DAP. The *V. cholerae* El Tor biotype strain C6706str2 was utilized in this study and mutant strains were generated using the pKAS32 suicide vector [67] using three fragments: 500 bp of sequence upstream of the gene of interest, 500 bp of sequence downstream of the gene of interest and cloned into the KpnI and SacI restriction sites of pKAS32 using by Gibson Assembly (NEB). P_{tac} inducible expression vectors were constructed by Gibson Assembly with inserts amplified by PCR and pEVS143 [68] or pMMB67EH [69] each linearized by EcoRI and BamHI, as well as pET28b digested with NcoI and XhoI. pEVS141 [70] is used as an empty vector control for experiments using pEVS143 derived constructs. Site-directed mutagenesis was performed using the SPRINP method [71]. Plasmids were introduced into *V. cholerae* through biparental conjugation using an *E. coli* BW29427 donor. Transformation of *E. coli* for ectopic expression experiments was performed using electroporation with DH10b for expression of pEVS143 and pMMB67EH derived plasmids and BL21(DE3) for pET28b based constructs.

Correlogy Bioinformatics Analysis

Our Correlogy software package is built on Kim and Price's approach [32] to calculate genetic co-occurrence. The source code, documentation, and a Docker container for this Python3 package are available at https://github.com/clinte14/correlogy. While VSP-1 is used to simplify the description of the method detailed below, both VSP-1 and 2 were independently analyzed in the same fashion. To establish maximum related subnetworks (MRS) for the genomic region of the VSP-1 island, a BLASTP amino acid sequence was performed to search

for each VSP-1 gene against the NCBI non-redundant protein database with an E-value cutoff of 10^{-4}. The BLAST results were limited to bacterial genomes, and all taxa belonging to the genus *Vibrio* were removed to avoid bias from closely related vertical inheritance. The BLAST results were used to generate a presence or absence matrix of VSP-1 homologues with all species along one axis and VSP-1 genes along the other axis. Next, a pairwise Pearson correlation value was calculated between all VSP-1 genes *i* and *j* using binary data from the above-mentioned presence/absence matrix*:*

$$r_{ij} = \frac{C_{ij}N - E_iE_j}{\sqrt{E_iE_j(N - E_i)(N - E_j)}},$$

where N is the total number of unique species returned from the BLAST search and C_{ij} the number of species with co-occurrence of genes *i and j.* While a Pearson correlation is warranted for a normally distributed binary data set, it does not account for indirect correlation. For example, if genes *i* and *j* individually associate with a third gene, a Pearson correlation will incorrectly calculate a correlation between *i* and *j.* To help correct for indirect correlation we calculate a partial correlation w_{ij} from the Pearson r_{ij}:

$$w_{ij} = \frac{P_{ij}}{\sqrt{P_{ii}P_{ij}}},$$

where the (*i, j*) element of the inverse matrix of Pearson r_{ij} is P_{ij} [32].

The partial correlation correction w_{ij} has the advantage of generating a normalized output that ranges between -1 to 1. For example, a w_{ij} of -1 reveals genes *i* and *j* never occur in the same species, while a value of 1 demonstrates genes *i* and *j* always co-occur in the same species. A w_{ij} of 0 is the amount of co-occurrence expected between unrelated genes *i* and *j* drawn from a normal distribution. Using the above-mentioned approach, a partial correlation value w_{ij} was calculated for all genes *i* to *j* in VSP-1 and VSP-2 (Supplemental Files 1 and 2). The single highest w_{ij} value for each VSP-1 gene was represented as an edge (i.e., line) in our

visualization (Fig. 1B, S1A, and S1B). Any set of genes that contains no further edges were assigned to a unique MRS that suggests functional association of the gene products within a unique gene network.

Genomic Identification, Structural, and Sequence Analyses of DcdV & DifV Homologs

DcdV from *V. cholerae* El Tor N16961 (WP_001901328.1) was identified as locus tag *vc0175*. DcdV and homologs profiles are performed using translated BLAST tblastn and run against the Nucleotide collection (nr/nt) database of National Center for Biotechnology Information (NCBI), using >40% similarities cutoff. For previously annotated domains, the Pfam feature in KEGG [72, 73] were utilized as a guide to determine DcdV homologs. Out of all the DcdV homologs, DcdV homologs from *Vibrio parahaemolyticus* O1: Kuk str. FDA_R31 (WP_020839904.1), *Proteus mirabilis* AR_379 (WP_108717204.1), and *E. coli* O78:H11 H10407 (ETEC) (WP_096882215.1) were analyzed in this study. Genomic contextual information from prokaryotic gene neighborhoods was retrieved from NCBI genome graphics feature to uncover *difV*-like gene, encoded as a hypothetical ORF 5' of the *dcdV* locus. If unannotated, the ORFinder feature from NCBI was used to determine the location and size of the putative *difV* locus. To predict the structure of DcdV from *V. cholerae,* the amino acid sequence was submitted to Phyre2 [39] and structural visualization was performed using PyMol (https://pymol.org). The amino acid and nucleotide alignments were analyzed using ClustalW Omega from EMBL-EBI web services [74] and LocARNA [75], respectively.

Identification and Characterization of Protein Homologs

Homology searches: To ensure the identification of a comprehensive set of homologs (close and remote), we started with six representative DcdV proteins across proteobacteria from *V. cholerae*, *V. parahaemolyticus*, *P. mirabilis*, and *E. coli* described above along with *E. cloacae* (WP_129996984.1), and *A. veronii* (WP_043825948.1) and performed homolog

searches using DELTABLAST [76] against all sequenced genomes across the tree of life in the NCBI RefSeq database [77–79]. Homology searches were conducted for each protein and the search results were aggregated; the numbers of homologs per species and of genomes carrying each of the query proteins were recorded. These proteins were clustered into orthologous families using the similarity-based clustering program BLASTCLUST [76].

Characterizing homologous proteins: Phyre2, InterProScan, HHPred, SignalP, TMHMM, Phobius, Pfam, and custom profile databases [39–41, 80–85] were used to identify signal peptides, transmembrane (TM) regions, known domains, and secondary structures of proteins in every genome. Custom scripts were written to consolidate the results [86–91], and the domain architectures and protein function predictions were visualized using the MolEvolvR web-app (http://jravilab.org/molevolvr/).

Phylogenetic analysis (MSA and Tree): Thousands of homologs from all six starting points for DcdV proteins were consolidated and representatives were chosen from distinct Lineages and Genera, containing both the N- and C-terminal DcdV domains (PLK and DCD domains). Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign [89] and MUSCLE [92, 93] (msa R package [94]). The phylogenetic trees were constructed using FastTree [95] FigTree [96] and the R package, ape [97].

Growth Curve Assays

Overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG in a 96-well microplate (Costar®). Growth was monitored by measuring OD_{600} every 15 minutes for 15 hour (h) using a BioTek plate reader with continuous, linear shaking.

Fluorescence Microscopy and Analysis

Cells were imaged as previously described [34]. Briefly, overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG. Cultures were grown and induced for 7-

8 h, at which point cells were diluted to an OD_{600} of 0.5 in 1X PBS, then membrane stain FM4-64 dye (ThermoFisher Scientific) was added to a final concentration of 20 µg/mL. 1% agarose pads in deionized water were cut into squares of approximately 20 x 20 mm and placed on microscope slides. 2 µl of diluted cultures were spotted onto a glass coverslip and then gently placed onto the agarose pad. FM4-64 signal was visualized using a Leica DM5000b epifluorescence microscope with a 100X-brightfield objective under RFP fluorescence channel. Images were captured using a Spot Pursuit CCD camera and an X-cite 120 Illumination system. Each slide was imaged with at least 20 fields of view for each biological replicate. Cell lengths were processed using the Fiji plugin MicrobeJ [98, 99], and data were visualized and analyzed using R [90] by quantifying the length of the curvilinear (medial) axis of detected cells.

Construction and screening of mutant gene libraries

DifV-insensitive DcdV constructs were generated by error-prone PCR (epPCR) using pDcdV (pCMW204) as the template. Three different concentrations of $MnCl_2$ (12.5 mM, 1.25 mM, and 125 µM) were used in triplicate using Taq polymerase (Invitrogen) and reactions containing the same $MnCl_2$ concentration were pooled. The PCR products were purified, using The Wizard® SV Gel *and* PCR Clean-Up Kit (Promega), and ligated to pEVS143 via Gibson Assembly. The assembled reactions were electroporated to *E. coli* DH10b and plasmid libraries were collected from ~ 30,000 representative colonies for each $MnCl_2$ concentration. Plasmid libraries were harvested using the Wizard® Plus SV Minipreps DNA purification Kit (Promega). Plasmid libraries were subsequently electroporated to *E. coli* BW29427 which were again plated and pooled to contain ~ 30,000 representative colonies. The *E. coli* BW29427 random mutant pDcdV libraries were conjugated with Δ*dcdV V. cholerae* on LB agar plates for 8 h, harvested, diluted, and spread on LB agar plates containing 1 mM IPTG and antibiotics, and grown overnight. ~ 5,000 colonies were screened in each library and all colonies exhibiting a wrinkled and small colony morphology, indicative of cell filamentation, were isolated and filamentation

was confirmed by fluorescence microscopy. Mutant pDcdV plasmids recovered from cells exhibiting cell filamentation were sequenced by Sanger sequencing. Mutations were reintroduced individually into the WT pDcdV construct using SPRINP mutagenesis [71] and reevaluated using fluorescence microscopy to confirm the DcdV variant's ability to remain constitutively active in Δ*dcdV V. cholerae*.

RNA Isolation, qRT-PCR, and Co-transcription Analysis

RNA isolation and qRT-PCR analysis were carried out as previously described [100, 101]. Briefly, triplicate overnight cultures were subcultured 1:1000 in 10 mL LB and grown to three different OD_{600}: 0.2 (Early Exponential), 1.0 (Late Exponential), and 2.5 (Stationary). 1 mL of each replicate was pelleted, and RNA was extracted using TRIzol® reagent following the manufacturer's directions (Thermo Fischer Scientific). RNA quality and quantity were determined using a NanoDrop spectrophotometer (Thermo Fischer Scientific). 5 µg of purified RNA was treated with DNase (Turbo™ DNase, Thermo Fischer Scientific). cDNA synthesis was carried out using SuperScript™ III Reverse Transcriptase (Thermo Fischer Scientific). cDNA was diluted 1:64 into molecular biology grade water and amplification was quantified using 2x SYBR Green (Applied Biosystems™). For measuring gene expressions or determining *orI*/*ter* ratios, 25 µL reactions consisted of 5 µL each of 0.625 µM primers 1 and 2, 12.5 µL of 2X SYBR master mix, and 2.5 µL of template (0.78 ng/µL cDNA for gene expression and 0.25 ng/µL genomic DNA for *orI*/*ter*). qRT-PCR reactions were performed in technical duplicates for biological triplicate samples and included no reverse transcriptase reaction controls ("no RT") to monitor for contaminating genomic DNA in purified RNA samples. qRT-PCR reaction thermo profile was 95°C for 20 seconds (s) then 40 cycles of 95°C for 2 s and 60°C for 30 s in the QuantStudio 3 Real-Time PCR system (Applied Biosystems™). The *gyrA* gene was used as an endogenous control to calculate relative quantification (ΔC_t).

642 To determine the co-transcription of *difV* and *dcdV*, PCR amplification was performed in
643 25 µL volumes using Q5 polymerase (NEB), 0.5 µM each of the forward and reverse primers as
644 indicated, 0.2 mM dNTPs, and 3.5 µL of cDNA or no RT control templates (0.78 ng/µL) from
645 RNA purified from WT and Δig^{222} *V. cholerae* grown to late exponential-phase in biological
646 triplicate. The thermal profile was 98°C for 30 s, 30 cycles of 98°C for 10 s, 55 °C for 30 s, 72 °C
647 for 10 sec and one cycle of 72 °C for 2 min. PCR products were loaded on a 1% agarose gel
648 and stained with EZ-Vision® (VWR). Images were taken using the GelDoc system (Bio-Rad).
649
650 **In-vitro Nucleic Acid Deamination Assay**
651 **Cell Lysate Preparation:** Overnight cultures were subcultured 1:333 and grown to an
652 OD_{600} of ~0.5 - 1.0. Cultures were induced with 1 mM IPTG, supplemented with 100 µM $ZnSO_4$,
653 and grown for an additional 3 hr. Cell pellets from 100 mL of induced cultures were harvested in
654 two successive 15 min centrifugation steps at 4,000 x g and 4°C. Supernatants were decanted
655 and pellets were snap frozen in an ethanol and dry ice bath and stored at -80° C. Pellets were
656 thawed on ice and suspended in 2 mL of lysis buffer A (50 mM $NaPO_4$, pH 7.3, 300 mM NaCl, 2
657 mM β-mercaptoethanol, 20% glycerol and Roche cOmplete protease inhibitor (1 tablet per 10
658 mL)). 1 mL of cell suspension was transferred to a microcentrifuge tube and sonicated on ice
659 using a Branson 450 Digital Sonifier (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off).
660 Crude lysates were centrifuged at 15,000 x g for 10 min at 4°C and clarified lysates were
661 transferred to fresh microcentrifuge tubes on ice. Clarified lysates were normalized for total
662 protein to 1.9 mg/mL using Bradford reagents and a BSA standard. 26.5 µL reactions composed
663 of lysis buffer A, nucleic acid substrates, and 3.5 µL of normalized clarified lysates were
664 assembled in PCR strip tubes, mixed by gentle pipetting, and incubated at room temperature
665 (~23°C) for 60 minutes. NH_4Cl solutions at the indicated concentration were dissolved in lysis
666 buffer A and substituted for nucleic acid substrates as positive ammonium controls.

667 **Ammonium Detection:** The evolution of NH_4^+ by deamination of the nucleic acid
668 substrates was observed using a phenol-hypochlorite reaction to produce indophenol in a clear
669 96-well microtiter plate and modified from Dong et al. 2015 [102]. The work of Ngo et al. [103]
670 was considered when designing the lysis buffer so as not to interfere with the phenol-
671 hypochlorite reaction. 50 µL of Reagent A (composition below) was added to each well followed
672 by 20 µL of the completed in vitro deamination reaction described above. The phenol-
673 hypochlorite reaction was initiated by the addition and gentle mixing of 50 µL Reagent B
674 (composition below) to the wells. The reaction was incubated at 35°C for 30 min and the ABS_{630}
675 was measured using a plate reader.
676 *Reagent A* = 1:1 (v/v), 6% (w/v) sodium hydroxide (Sigma) in water: 1.5% (v/v) sodium
677 hypochlorite solution (Sigma, reagent grade) in water.
678 *Reagent B* = 1:1:0.04 (v/v/v), water: 0.5% (w/v) sodium nitroprusside (Sigma) in water:
679 phenol solution (Sigma, P4557)
680
681 **Western Blot**
682 Strains containing DcdV- and variant- C-terminal 6x-histidine fusions were grown,
683 induced, and harvested as described previously above (See In-vitro Nucleic Acid Deamination
684 Assay: Cell Lysate Prep), except for the His-tag fusion (pGBS98) which are induced for only 2 h
685 with 100 µM IPTG and not subjected to sonication. The cell pellets were resuspended in 2 mL of
686 chilled 1X PBS and subsequently normalized to OD of 1.0. 1 mL aliquots were collected by
687 centrifugation at 15k x g for 1 min. Cell pellets were subsequently resuspended in 90 µL of lysis
688 buffer A and 30 µL of 4x Laemmli buffer, denatured for 10 minutes at 65°C, and centrifuged at
689 15k x g for 10 minutes. 5 µL of samples were loaded into a precast 4-20% sodium dodecyl
690 sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (Mini-PROTEAN TGX Precast
691 Protein Gels, Bio-Rad) alongside size standards (Precision Protein Plus, Bio-Rad). Gels were
692 run at room temperature for 90 min at 100 V in 1x Tris/glycine/SDS running buffer. Proteins

693 were transferred to nitrocellulose membranes (Optitran). The membranes were blocked using
694 5% skim milk and incubated with 1:5000 THE™ His Tag Antibody, mAb, Mouse (GenScript)
695 followed by 1:4000 Goat Anti-Mouse IgG Antibody (H&L) [HRP], pAb (GenScript), treated with
696 Pierce™ ECL Western Blotting Substrate, and imaged using an Amersham™ Imager 600.
697
698 **UPLC-MS/MS Quantification of In Vitro and In Vivo Deoxynucleotides**
699 Deoxynucleotide concentrations were determined as previously described [104] with
700 minor modifications. For measuring in vivo intracellular deoxynucleotide concentrations,
701 overnight cultures were subcultured 1:1000 and grown to OD_{600} of ~1.0. Plasmid expression
702 was induced by the addition of 1 mM IPTG for 1 h, and 1 mL of cultures were collected by
703 centrifugation at 15,000 x g for 1 min. Cell pellets were resuspended in 200 µL of chilled
704 extraction buffer [acetonitrile, methanol, ultra-pure water, formic acid (2:2:1:0.02, v/v/v/v)]. To
705 normalize in vivo nucleotide samples, an additional cell pellet was collected from 1 mL of culture
706 by centrifugation at 15,000 x g for 1 min, resuspended in 200 µL lysis buffer B (20 mM Tris·HCl,
707 1% SDS, pH 6.8), and denatured for 10 minutes at 60°C. Denatured lysates were centrifuged at
708 15,000 x g for 1 min to pellet cellular debris, and the supernatant was used to quantify the total
709 protein concentration in the sample using the DC protein assay (Bio-Rad) a BSA standard curve
710 [34]. The concentrations of deoxynucleotides detected by UPLC-MS/MS were then normalized
711 to total protein in each sample.
712 For the quantification of deoxynucleotides in vitro *E. coli* BL21(DE3) clarified lysates
713 were prepared as described for the deamination experiment above and normalized to 20 mg/mL
714 of total protein and 200 µL of normalized clarified lysates were assembled in PCR strip tubes.
715 To measure abundance of dUMP and dUTP prior to the addition of 1 µM dCTP, 20 µL of
716 normalized clarified lysates were added to 200 µL of chilled extraction buffer. 20 µL of 10 µM
717 dCTP was then added to the remaining clarified lysates and 20 µL lysates aliquots were

718 removed 1, 5, 10, and 30 minutes after the addition of dCTP and mixed in 200 µL chilled
719 extraction buffer.
720 All samples resuspended in extraction buffer, in vivo and in vitro, were immediately
721 incubated at -20°C for 30 minutes after collection and centrifuged at 15,000 x g for 1 min. The
722 supernatant was transferred to a new tube, dried overnight in a speed vacuum, and finally
723 resuspended in 100 µL ultra-pure water. Experimental samples and deoxynucleotides standards
724 [1.9, 3.9, 7.8, 15.6, 31.3, 62.5, and 125 nM of dATP (Invitrogen), dGTP (Invitrogen), dTTP,
725 (Invitrogen), dCTP (Invitrogen), dCMP (Sigma), dUTP (Sigma), and dUMP (Sigma)] were
726 analyzed by UPLC-MS/MS using an Acquity Ultra Performance LC system (Waters) coupled
727 with a Xevo TQ-S mass spectrometer (Waters) with an ESI source in negative ion mode. The
728 MS parameters were as follows: capillary voltage, 1.0 kV; source temperature, 150°C;
729 desolvation temperature, 400°C; cone gas, 120 L/hr. Five microliter of each sample was
730 separated in reverse phase using Acquity UPLC Premier BEH C18, 2.1 x 100 mm, 1.7 µm
731 particle size, VanGuard FIT at a flow rate of 0.3 mL/min with the following gradient of solvent A
732 (8mM DMHA (N,N-dimethylhexylamine) + 2.8 mM acetic acid in water, pH~9) to solvent B
733 (methanol): *t* = 0 min; A-100%:B-0%, *t* = 10 min; A-60%:B-40%, *t* =10.5; A-100%:B-0%, *t* = 15
734 min; A-100%:B-0% (end of gradient). The conditions of the MRM transitions were as follows
735 [cone voltage (V), collision energy (eV)]: dATP, 490 > 159 (34, 34); dCTP, 466 > 159 (34, 34);
736 dGTP, 506 > 159 (15, 46); dTTP, 481 > 159 (25, 34); dUTP, 467 > 159 (25, 34); dCMP, 306 >
737 97 (43, 22); dUMP, 306 > 111 (22, 22).
738
739 **Phage Infection and Plaque Assays**
740 *V. cholerae* phages ICP1 and ICP3 were provided by Wai-Leung Ng at Tuft University
741 School of Medicine. ICP1 was propagated on *V. cholerae* E7946, while ICP3 were propagated
742 on *V. cholerae* C6706str2 in LB, and their titer was determined using the small drop plaque
743 assay method, as previously described [10]. Briefly, 1 ml of overnight cultures were mixed with 9

744 ml of MMB agar (LB + 0.1 mM MnCl2 + 5 mM MgCl2 + 5 mM CaCl2 + 0.5% agar), tenfold serial
745 dilutions of phages in MMB were dropped on top of them, and incubated overnight at 35°C. The
746 viral titer is expressed as plaque forming units per mL (pfu/mL). 4 mL of *V. cholerae* overnight
747 cultures were diluted 1:1000 in MMB medium. 145 µL of the diluted cultures, in three sets of
748 biological replicates, were transferred and incubated at 35°C in a 96-well microplate (Costar®).
749 Once the OD_{600} reached ~0.1, 5 µL of phages with a final MOI of 0.1 were added to each
750 biological replicate. Cultures were infected at room temperature (~23°C) for 12 h in a
751 SpectraMax M5 Plate Reader with continuous shaking and OD_{600} measurements taken every
752 2.5 min.
753 *Shigella flexneri* strain PE577 [54] cells transformed with the pVector (pMMB67eh) and
754 each of the associated pDcdV plasmids were grown in LB medium and incubated with aeration
755 at 37° C overnight. The following day, 20 µL of each of the overnight cultures were used to
756 inoculate fresh medium in a 96-well microtiter plate with a final volume of 200 µL/well.
757 Depending on the experimental condition, wells were supplemented with and without IPTG (100
758 µM final concentration) and/or phage Sf6 [55] at an MOI of 0.1 phage per cell. Initial cell
759 densities of the overnight cultures were experimentally determined by plating and found to be
760 within a factor of two of one another. For all experiments, three biological replicates were tested.
761 Additionally, the plates were set up with each unique condition having three technical replicates.
762 Plate reader assays were conducted using a Molecular Devices FilterMax F5 plate reader, as
763 previously described [105]. Briefly, the plates were incubated at 37°C for 6 h. Every five
764 minutes, the plate was mixed and aerated by orbital shaking before an absorbance (595 nm)
765 reading was taken. After the kinetic assay was complete an aliquot from each of the replicates
766 was removed and used to determine the endpoint titer via plaque assay.
767
768
769

Statistical Analysis

As specified in the figure legends, all of the statistical analyses for the violin plots were performed with R statistical computing software [90], while other data were analyzed in GraphPad Prism Software. Statistically significances denote as the following: a single asterisk (*) indicates $p < 0.05$; double asterisks (**) indicate $p < 0.01$; triple asterisks (***) indicate $p < 0.001$; and quadruple asterisks (****) indicate $p < 0.0001$. Means ± SEM and specific n values are reported in each figure legend.

REFERENCE

1. Thompson FL, Iida T, Swings J (2004) Biodiversity of *Vibrios*. Microbiol Mol Biol Rev 68:403–431 . https://doi.org/10.1128/MMBR.68.3.403-431.2004
2. Son MS, Taylor RK (2011) Genetic Screens and Biochemical Assays to Characterize *Vibrio cholerae* O1 Biotypes: Classical and El Tor. Curr Protoc Microbiol 22:6A.2.1-6A.2.17 . https://doi.org/10.1002/9780471729259.mc06a02s22
3. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ (2002) Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci 99:1556–1561 . https://doi.org/10.1073/pnas.042667999
4. Hu D, Liu B, Feng L, Ding P, Guo X, Wang M, Cao B, Reeves PR, Wang L (2016) Origins of the current seventh cholera pandemic. Proc Natl Acad Sci 113:E7730–E7739 . https://doi.org/10.1073/pnas.1608732113
5. O'Shea YA, Finnan S, Reen FJ, Morrissey JP, O'Gara F, Boyd EF (2004) The Vibrio seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in V. vulnificus. Microbiology 150:4053–4063 . https://doi.org/10.1099/mic.0.27172-0

6. Murphy RA, Boyd EF (2008) Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. J Bacteriol 190:636–647 . https://doi.org/10.1128/JB.00562-07

7. Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, Lanata CF, Hall E, Miranda H, Huapaya B, Vargas G. C, Luna MA, Sack DA, Yamasaki S, Nair GB (2009) Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. J Med Microbiol 58:342–354 . https://doi.org/10.1099/jmm.0.005397-0

8. Beyhan S, Tischler AD, Camilli A, Yildiz FH (2006) Differences in Gene Expression between the Classical and El Tor Biotypes of *Vibrio cholerae* O1. Infect Immun 74:3633–3642 . https://doi.org/10.1128/IAI.01750-05

9. Davies BW, Bogard RW, Young TS, Mekalanos JJ (2012) Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for *V. cholerae* Virulence. Cell 149:358–370 . https://doi.org/10.1016/j.cell.2012.01.053

10. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574:691–695 . https://doi.org/10.1038/s41586-019-1605-5

11. Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger A-K, Khataokar A, O'Hara BJ, Behrmann L V., Neiditch MB, Benning C, Waters CM, Ng W-L (2018) Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. Proc Natl Acad Sci 115:E6048–E6055 . https://doi.org/10.1073/pnas.1801233115

12. O'Donovan GA, Neuhard J (1970) Pyrimidine metabolism in microorganisms. Bacteriol Rev 34:278–343 . https://doi.org/10.1128/MMBR.34.3.278-343.1970

13. Weiss B (2007) The Deoxycytidine Pathway for Thymidylate Synthesis in *Escherichia coli*. J Bacteriol 189:7922–7926 . https://doi.org/10.1128/JB.00461-07

14. Weiss B, Wang L (1994) De novo synthesis of thymidylate via deoxycytidine in dcd

(dCTP deaminase) mutants of *Escherichia coli*. J Bacteriol 176:2194–2199 . https://doi.org/10.1128/JB.176.8.2194-2199.1994

15. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L (2018) Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. Proc Natl Acad Sci 115:E3201–E3210 . https://doi.org/10.1073/pnas.1720897115

16. Chiu Y-L, Greene WC (2008) The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements. Annu Rev Immunol 26:317–353 . https://doi.org/10.1146/annurev.immunol.26.021607.090350

17. Lada AG, Frahm Krick C, Kozmin SG, Mayorov VI, Karpova TS, Rogozin IB, Pavlov YI (2011) Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochem 76:131–146 . https://doi.org/10.1134/S0006297911010135

18. Stavrou S, Ross SR (2015) APOBEC3 Proteins in Viral Immunity. J Immunol 195:4565–70 . https://doi.org/10.4049/jimmunol.1501504

19. Sharma S, Patnaik SK, Thomas Taggart R, Kannisto ED, Enriquez SM, Gollnick P, Baysal BE (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun 6:6881 . https://doi.org/10.1038/ncomms7881

20. Kreuzer KN, Brister JR (2010) Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J 7:358 . https://doi.org/10.1186/1743-422X-7-358

21. You L, Suthers PF, Yin J (2002) Effects of Escherichia coli Physiology on Growth of Phage T7 In Vivo and In Silico. J Bacteriol 184:1888–1894 . https://doi.org/10.1128/JB.184.7.1888-1894.2002

22. Endy D, Kong D, Yin J (1997) Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. Biotechnol Bioeng 55:375–389 . https://doi.org/10.1002/(SICI)1097-0290(19970720)55:2<375::AID-BIT15>3.0.CO;2-G

847 23. Zhang Y, Maley F, Maley GF, Duncan G, Dunigan DD, Van Etten JL (2007)
848 Chloroviruses Encode a Bifunctional dCMP-dCTP Deaminase That Produces Two Key
849 Intermediates in dTTP Formation. J Virol 81:7662–7671 .
850 https://doi.org/10.1128/jvi.00186-07
851 24. Graziani S, Xia Y, Gurnon JR, Van Etten JL, Leduc D, Skouloubris S, Myllykallio H, Liebl
852 U (2004) Functional Analysis of FAD-dependent Thymidylate Synthase ThyX from
853 *Paramecium bursaria* Chlorella Virus-1. J Biol Chem 279:54340–54347 .
854 https://doi.org/10.1074/jbc.M409121200
855 25. Zhang Y, Moriyama H, Homma K, Van Etten JL (2005) Chlorella Virus-Encoded
856 Deoxyuridine Triphosphatases Exhibit Different Temperature Optima. J Virol 79:9945–
857 9953 . https://doi.org/10.1128/JVI.79.15.9945-9953.2005
858 26. Chiu CS, Ruettinger T, Flanegan JB, Greenberg GR (1977) Role of deoxycytidylate
859 deaminase in deoxyribonucleotide synthesis in bacteriophage T4 DNA replication. J Biol
860 Chem 252:8603–8608 . https://doi.org/10.1016/S0021-9258(19)75263-6
861 27. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W (2003) Bacteriophage T4
862 Genome. Microbiol Mol Biol Rev 67:86–156 . https://doi.org/10.1128/MMBR.67.1.86-
863 156.2003
864 28. Guzmán EC, Martín CM (2015) Thymineless death, at the origin. Front Microbiol 6:1–7 .
865 https://doi.org/10.3389/fmicb.2015.00499
866 29. Itsko M, Schaaper RM (2014) dGTP Starvation in *Escherichia coli* Provides New Insights
867 into the Thymineless-Death Phenomenon. PLoS Genet 10:e1004310 .
868 https://doi.org/10.1371/journal.pgen.1004310
869 30. Kuong KJ, Kuzminov A (2010) Stalled replication fork repair and misrepair during
870 thymineless death in *Escherichia coli*. Genes to Cells 15:619–634 .
871 https://doi.org/10.1111/j.1365-2443.2010.01405.x
872 31. Ahmad SI, Kirk SH, Eisenstark A (1998) Thymine Metabolism and Thymineless Death in

Prokaryotes and Eukaryotes. Annu Rev Microbiol 52:591–625 . https://doi.org/10.1146/annurev.micro.52.1.591

32. Kim PJ, Price ND (2011) Genetic co-occurrence network across sequenced microbes. PLoS Comput Biol 7: . https://doi.org/10.1371/journal.pcbi.1002340

33. Bartlett TM, Bratton BP, Duvshani A, Miguel A, Sheng Y, Martin NR, Nguyen JP, Persat A, Desmarais SM, VanNieuwenhze MS, Huang KC, Zhu J, Shaevitz JW, Gitai Z (2017) A Periplasmic Polymer Curves *Vibrio cholerae* and Promotes Pathogenesis. Cell 168:172-185.e15 . https://doi.org/10.1016/j.cell.2016.12.019

34. Fernandez NL, Hsueh BY, Nhu NTQ, Franklin JL, Dufour YS, Waters CM (2020) *Vibrio cholerae* adapts to sessile and motile lifestyles by cyclic di-GMP regulation of cell shape. Proc Natl Acad Sci 117:29046–29054 . https://doi.org/10.1073/pnas.2010199117

35. Raabe CA, Hoe CH, Randau G, Brosius J, Tang TH, Rozhdestvensky TS (2011) The rocks and shallows of deep RNA sequencing: Examples in the *Vibrio cholerae* RNome. RNA 17:1357–1366 . https://doi.org/10.1261/rna.2682311

36. Papenfort K, Förstner KU, Cong J-P, Sharma CM, Bassler BL (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. Proc Natl Acad Sci 112:E766–E775 . https://doi.org/10.1073/pnas.1500203112

37. Krin E, Pierlé SA, Sismeiro O, Jagla B, Dillies M-AA, Varet H, Irazoki O, Campoy S, Rouy Z, Cruveiller S, Médigue C, Coppée J-YY, Mazel D (2018) Expansion of the SOS regulon of *Vibrio cholerae* through extensive transcriptome analysis and experimental validation. BMC Genomics 19:1–18 . https://doi.org/10.1186/s12864-018-4716-8

38. Short FL, Akusobi C, Broadhurst WR, Salmond GPC (2018) The bacterial Type III toxin-antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent antiviral abortive infection activity. Sci Rep 8:1013 . https://doi.org/10.1038/s41598-017-18696-x

39. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web

portal for protein modeling, prediction and analysis. Nat Protoc 10:845–58 . https://doi.org/10.1038/nprot.2015.053

40. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240 . https://doi.org/10.1093/bioinformatics/btu031

41. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL (2017) InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res 45:D190–D199 . https://doi.org/10.1093/nar/gkw1107

42. Leipe DD, Koonin E V., Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. J Mol Biol 333:781–815 . https://doi.org/10.1016/j.jmb.2003.08.040

43. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1:945–51 . https://doi.org/10.1002/j.1460-2075.1982.tb01276.x

44. Khan SR, Kuzminov A (2019) Thymineless Death in *Escherichia coli* Is Unaffected by Chromosomal Replication Complexity. J Bacteriol 201: . https://doi.org/10.1128/JB.00797-18

45. Hou H-F, Liang Y-H, Li L-F, Su X-D, Dong Y-H (2008) Crystal structures of *Streptococcus mutans* 2'-deoxycytidylate deaminase and its complex with substrate analog and

allosteric regulator dCTP x Mg2+. J Mol Biol 377:220–31 . https://doi.org/10.1016/j.jmb.2007.12.064

46. Marx A, Alian A (2015) The first crystal structure of a dTTP-bound deoxycytidylate deaminase validates and details the allosteric-inhibitor binding site. J Biol Chem 290:682–90 . https://doi.org/10.1074/jbc.M114.617720

47. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M, Schwede F, Mekalanos JJ, Shao S, Lee ASY, Kranzusch PJ (2020) CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell 182:38-49.e17 . https://doi.org/10.1016/j.cell.2020.05.019

48. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC Family of Polynucleotide (Deoxy)cytidine Deaminases. Mol Biol Evol 22:367–377 . https://doi.org/10.1093/molbev/msi026

49. Seed KD, Bodi KL, Kropinski AM, Ackermann H-W, Calderwood SB, Qadri F, Camilli A (2011) Evidence of a Dominant Lineage of *Vibrio cholerae*-Specific Lytic Bacteriophages Shed by Cholera Patients over a 10-Year Period in Dhaka, Bangladesh. MBio 2:e00334-10 . https://doi.org/10.1128/mBio.00334-10

50. Yen M, Cairns LS, Camilli A (2017) A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. Nat Commun 8:14187 . https://doi.org/10.1038/ncomms14187

51. O'Hara BJ, Barth ZK, McKitterick AC, Seed KD (2017) A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. PLoS Genet 13:e1006838 . https://doi.org/10.1371/journal.pgen.1006838

52. McKitterick AC, Seed KD (2018) Anti-phage islands force their target phage to directly mediate island excision and spread. Nat Commun 9:2348 . https://doi.org/10.1038/s41467-018-04786-5

53. Hays SG, Seed KD (2020) Dominant *Vibrio cholerae* phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. Elife 9: . https://doi.org/10.7554/eLife.53200

54. Morona R, Mavris M, Fallarino A, Manning PA (1994) Characterization of the rfc region of *Shigella flexneri*. J Bacteriol 176:733–747 . https://doi.org/10.1128/JB.176.3.733-747.1994

55. Casjens SR (2004) The Chromosome of *Shigella flexneri* Bacteriophage Sf6: Complete Nucleotide Sequence, Genetic Mosaicism, and DNA Packaging*1. J Mol Biol 339:379–394 . https://doi.org/10.1016/S0022-2836(04)00377-8

56. Jemielita M, Wingreen NS, Bassler BL Quorum sensing controls *Vibrio cholerae* multicellular aggregate formation. Elife. https://doi.org/10.7554/eLife.42057

57. Johansson E, Fanø M, Bynck JH, Neuhard J, Larsen S, Sigurskjold BW, Christensen U, Willemoës M (2005) Structures of dCTP Deaminase from *Escherichia coli* with Bound Substrate and Product. J Biol Chem 280:3051–3059 . https://doi.org/10.1074/jbc.M409534200

58. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, Rosenberg MM, Tal N, Melamed S, Amitai G, Sorek R (2020) Prokaryotic viperins produce diverse antiviral molecules. Nature. https://doi.org/10.1038/s41586-020-2762-2

59. Ayinde D, Casartelli N, Schwartz O (2012) Restricting HIV the SAMHD1 way: Through nucleotide starvation. Nat Rev Microbiol 10:675–680 . https://doi.org/10.1038/nrmicro2862

60. Shepard C, Xu J, Holler J, Kim DH, Mansky LM, Schinazi RF, Kim B (2019) Effect of induced dNTP pool imbalance on HIV-1 reverse transcription in macrophages. Retrovirology 16:29 . https://doi.org/10.1186/s12977-019-0491-0

61. Oo A, Kim DH, Schinazi RF, Kim B (2020) Viral protein X reduces the incorporation of mutagenic noncanonical rNTPs during lentivirus reverse transcription in macrophages. J

977 Biol Chem 295:657–666 . https://doi.org/10.1074/jbc.RA119.011466

978 62. Harms A, Brodersen DE, Mitarai N, Gerdes K (2018) Toxins, Targets, and Triggers: An
979 Overview of Toxin-Antitoxin Biology. Mol Cell 70:768–784 .
980 https://doi.org/10.1016/j.molcel.2018.01.003

981 63. Makarova KS, Wolf YI, Snir S, Koonin E V. (2011) Defense Islands in Bacterial and
982 Archaeal Genomes and Prediction of Novel Defense Systems. J Bacteriol 193:6039–
983 6056 . https://doi.org/10.1128/JB.05535-11

984 64. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R (2018)
985 Systematic discovery of antiphage defense systems in the microbial pangenome. Science
986 (80-) 359:eaar4120 . https://doi.org/10.1126/science.aar4120

987 65. Bernheim A, Sorek R (2020) The pan-immune system of bacteria: antiviral defence as a
988 community resource. Nat Rev Microbiol 18:113–119 . https://doi.org/10.1038/s41579-
989 019-0278-2

990 66. Westra ER, Swarts DC, Staals RHJ, Jore MM, Brouns SJJ, van der Oost J (2012) The
991 CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. Annu
992 Rev Genet 46:311–339 . https://doi.org/10.1146/annurev-genet-110711-155447

993 67. Skorupski K, Taylor RK (1996) Positive selection vectors for allelic exchange. Gene
994 169:47–52 . https://doi.org/10.1016/0378-1119(95)00793-8

995 68. Bose JL, Rosenberg CS, Stabb E V. (2008) Effects of luxCDABEG induction in *Vibrio*
996 *fischeri*: Enhancement of symbiotic colonization and conditional attenuation of growth in
997 culture. Arch Microbiol 190:169–183 . https://doi.org/10.1007/s00203-008-0387-1

998 69. Fürste JP, Pansegrau W, Frank R, Blöcker H, Scholz P, Bagdasarian M, Lanka E (1986)
999 Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP
1000 expression vector. Gene 48:119–131 . https://doi.org/10.1016/0378-1119(86)90358-6

1001 70. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb E V (2006) and pES213-Derived Tools
1002 for Analyzing Symbiotic. Society 72:802–810 . https://doi.org/10.1128/AEM.72.1.802

71. Edelheit O, Hanukoglu A, Hanukoglu I (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. BMC Biotechnol 9:61 . https://doi.org/10.1186/1472-6750-9-61

72. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30 . https://doi.org/10.1093/nar/28.1.27

73. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M (2019) New approach for understanding genome variations in KEGG. Nucleic Acids Res 47:D590–D595 . https://doi.org/10.1093/nar/gky962

74. Madeira F, Madhusoodanan N, Lee J, Tivey ARN, Lopez R (2019) Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. Curr Protoc Bioinforma 66:e74 . https://doi.org/10.1002/cpbi.74

75. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA 18:900–14 . https://doi.org/10.1261/rna.029041.111

76. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421 . https://doi.org/10.1186/1471-2105-10-421

77. Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, Tolstoy I, Zaslavsky L (2015) Update on RefSeq microbial genomes resources. Nucleic Acids Res 43:D599–D605 . https://doi.org/10.1093/nar/gku1062

78. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C,

Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-45 . https://doi.org/10.1093/nar/gkv1189

79. Federhen S (2012) The NCBI Taxonomy database. Nucleic Acids Res 40:D136–D143 . https://doi.org/10.1093/nar/gkr1178

80. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248 . https://doi.org/10.1093/nar/gki408

81. Nielsen H (2017) Predicting Secretory Proteins with SignalP. Methods Mol Biol 1611:59–73 . https://doi.org/10.1007/978-1-4939-7015-5_6

82. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. J Mol Biol 305:567–580 . https://doi.org/10.1006/jmbi.2000.4315

83. Käll L, Krogh A, Sonnhammer ELL (2004) A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol 338:1027–1036 . https://doi.org/10.1016/j.jmb.2004.03.016

84. Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res 35: . https://doi.org/10.1093/nar/gkm256

85. Mistry J, Finn R (2007) Pfam. Methods Mol Biol 396:43–58 . https://doi.org/10.1007/978-1-59745-515-2_4

86. Wilkins D (2020) gggenes. available: https://github.com/wilkox/gggenes

87. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H

(2019) Welcome to the Tidyverse. J Open Source Softw 4:1686 . https://doi.org/10.21105/joss.01686

88. Müller K (2017) here: A Simpler Way to Find Your Files. available: https://cran.r-project.org/web/packages/here/index.html

89. Lassmann T (2019) Kalign 3: multiple sequence alignment of large datasets. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz795

90. R Core Team (2019) R: A language and environment for statistical computing. In: R Found. Stat. Comput. available: https://www.r-project.org/

91. RStudio Team (2020) RStudio: Integrated Development Environment for R. available: https://www.rstudio.com/

92. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. https://doi.org/10.1093/nar/gkh340

93. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-5-113

94. Bonatesta E, Horejs-Kainrath C, Bodenhofer U (2020) Multiple Sequence Alignment. available: https://www.bioinf.jku.at/software/msa/

95. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490 . https://doi.org/10.1371/journal.pone.0009490

96. Rambaut A (2009) FigTree. availabe: https://tree.bio.ed.ac.uk/software/figtree/

97. Paradis E, Blomberg S, Bolker B, Brown J, Claramunt S, Claude J, Cuong HS, Desper R, Didier G, Durand B, Dutheil J, Ewing R, Gascuel O, Guillerme T, Heibl C, Ives A, Jone B, Krah F, Lawson D, Lefort V, Legendre P, Lemon J, Louvel G, Marcon E, McCloskey R, Nylander J, Opgen-Rhein R, Popescu A-A, Royer-Carenzi M, Schliep K, Strimmer K, de Vienne D (2020) ape: Analyses of Phylogenetics and Evolution. available: https://ape-package.ird.fr/

98. Ducret A, Quardokus EM, Brun Y V. (2016) MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol 1:16077 . https://doi.org/10.1038/nmicrobiol.2016.77

99. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676–682 . https://doi.org/10.1038/nmeth.2019

100. Fernandez NL, Waters CM (2019) Cyclic di-GMP increases catalase production and hydrogen peroxide tolerance in *Vibrio cholerae*. Appl Environ Microbiol 85: . https://doi.org/10.1128/AEM.01043-19

101. Dryselius R, Izutsu K, Honda T, Iida T (2008) Differential replication dynamics for large and small *Vibrio* chromosomes affect gene dosage, expression and location. BMC Genomics 9:559 . https://doi.org/10.1186/1471-2164-9-559

102. Dong H, Liu Y, Zu X, Li N, Li F, Zhang D (2015) An enzymatic assay for high-throughput screening of cytidine-producing microbial strains. PLoS One 10:e0121612 . https://doi.org/10.1371/journal.pone.0121612

103. Ngo TT, Phan APH, Yam CF, Lenhoff HM (1982) Interference in determination of ammonia with the hypochlorite-alkaline phenol method of Berthelot. Anal Chem 54:46–49 . https://doi.org/10.1021/ac00238a015

104. Massie JP, Reynolds EL, Koestler BJ, Cong J-P, Agostoni M, Waters CM (2012) Quantification of high-specificity cyclic diguanylate signaling. Proc Natl Acad Sci 109:12746–12751 . https://doi.org/10.1073/pnas.1115663109

105. Dover JA, Burmeister AR, Molineux IJ, Parent KN (2016) Evolved populations of *Shigella flexneri* phage Sf6 acquire large deletions, altered genomic architecture, and faster life cycles. Genome Biol Evol 8:2827–2840 . https://doi.org/10.1093/gbe/evw177

106. Evans DJ, Evans DG (1973) Three Characteristics Associated with Enterotoxigenic

Escherichia coli Isolated from Man. Infect Immun 8:322–328 . https://doi.org/10.1128/IAI.8.3.322-328.1973

107. Helene Thelin K, Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. Infect Immun 64:2853–2856 . https://doi.org/10.1128/iai.64.7.2853-2856.1996

108. Miller VL, DiRita VJ, Mekalanos JJ (1989) Identification of toxS, a regulatory gene whose product enhances ToxR-mediated activation of the cholera toxin promoter. J Bacteriol 171:1288–1293 . https://doi.org/10.1128/jb.171.3.1288-1293.1989

109. Lüdeke CHM, Kong N, Weimer BC, Fischer M, Jones JL (2015) Complete Genome Sequences of a Clinical Isolate and an Environmental Isolate of *Vibrio parahaemolyticus*. Genome Announc 3:e00216-15 . https://doi.org/10.1128/genomeA.00216-15

110. Lutgring JD, Machado M-J, Benahmed FH, Conville P, Shawar RM, Patel J, Brown AC (2017) FDA-CDC Antimicrobial Resistance Isolate Bank: a Publicly Available Resource To Support Research, Development, and Regulatory Requirements. J Clin Microbiol 56: . https://doi.org/10.1128/JCM.01415-17

111. Parent KN, Erb ML, Cardone G, Nguyen K, Gilcrease EB, Porcek NB, Pogliano J, Baker TS, Casjens SR (2014) OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in *Shigella*. Mol Microbiol 92:47–60 . https://doi.org/10.1111/mmi.12536

112. Vanbleu E, Marchal K, Vanderleyden J (2004) Genetic and physical map of the pLAFR1 vector. DNA Seq - J DNA Seq Mapp 15:225–227 . https://doi.org/10.1080/10425170410001723949

FIGURE LEGENDS-located with the figures

Fig. 1

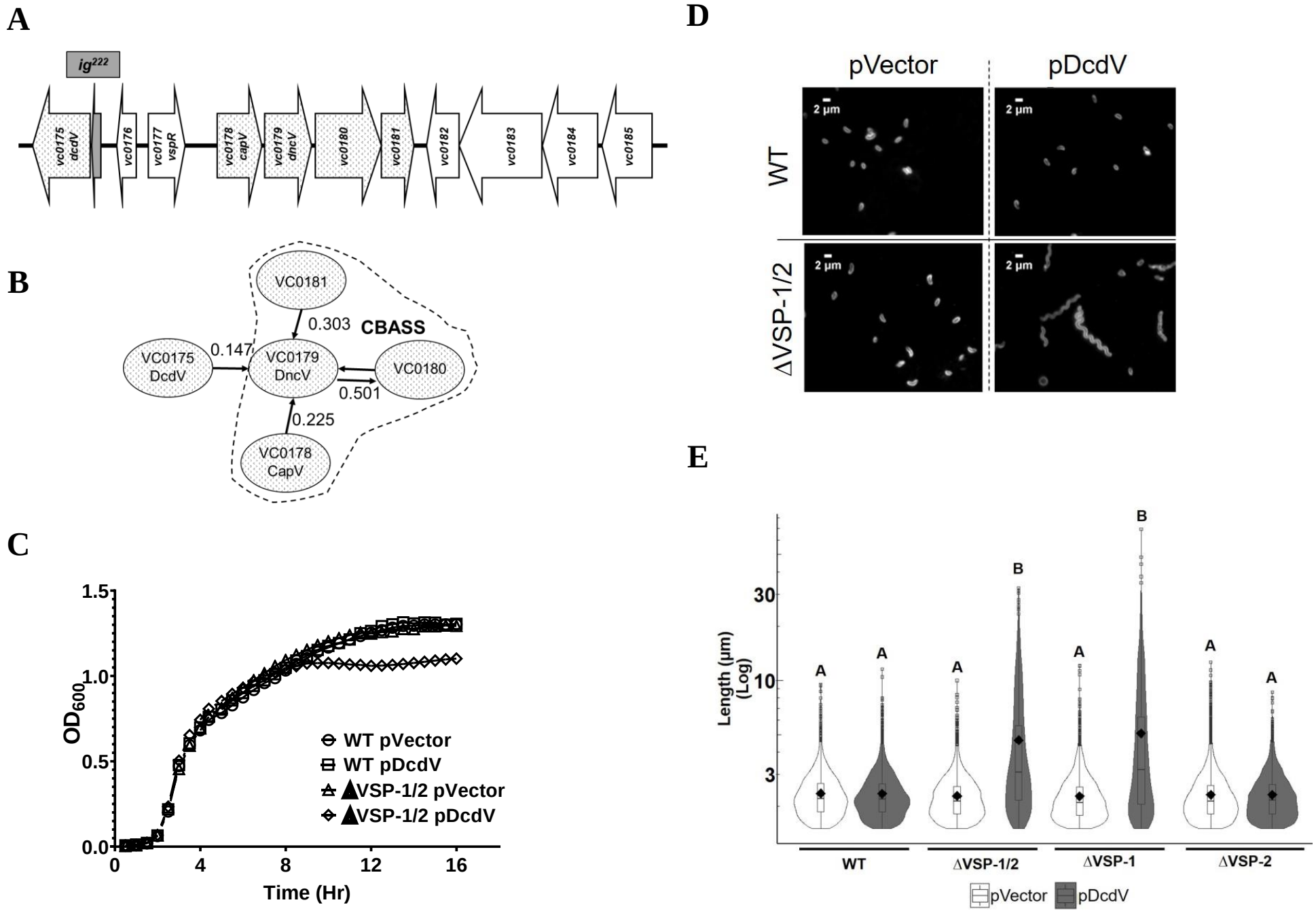
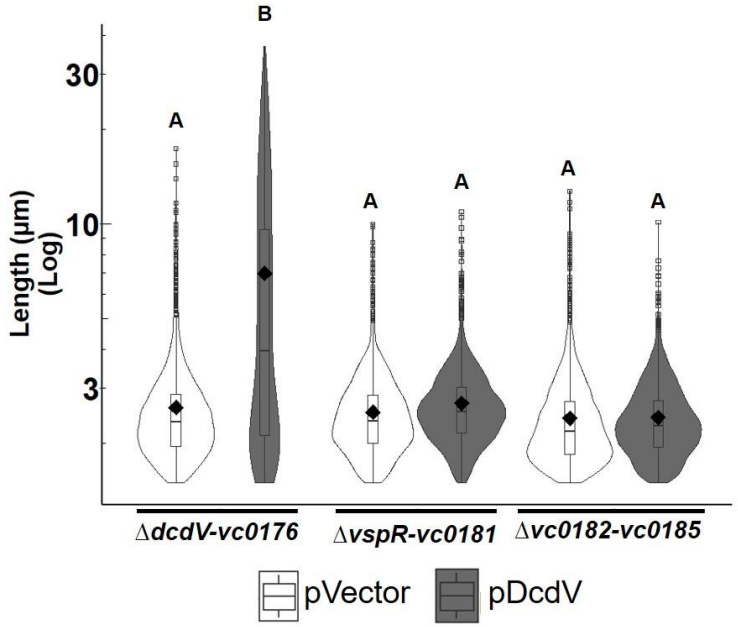

A
ig222
vc0175 dcdV
vc0176
vc0177 vspR
vc0178 capV
vc0179 dncV
vc0180
vc0181
vc0182
vc0183
vc0184
vc0185
B
VC0181
0.303
CBASS
VC0175 DcdV
0.147
VC0179 DncV
VC0180
0.501
0.225
VC0178 CapV
C
1.5
1.0
0.5
0.0
OD600
0
4
8
12
16
Time (Hr)
WT pVector
WT pDcdV
ΔVSP-1/2 pVector
ΔVSP-1/2 pDcdV
D
pVector
pDcdV
WT
ΔVSP-1/2
2 μm
E
Length (μm) (Log)
30
10
3
A
B
WT
ΔVSP-1/2
ΔVSP-1
ΔVSP-2
pVector
pDcdV

Fig. 1: DcdV promotes filamentation in *V. cholerae* in the absence of VSP-1

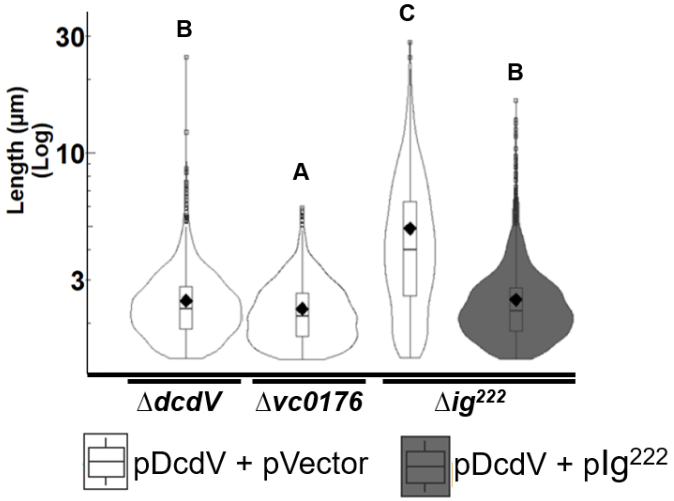
(**A**) Cartoon schematic of VSP-1 and (**B**) the Correlogy gene network prediction for *dncV* where arrows show the highest partial correlation W_{ij} each individual VSP-1 gene has to another. (**C**) Growth of WT *V. cholerae* and ΔVSP-1/2 strains with the vector or pDcdV. Data represent the mean ± SEM, *n*=3. (**D**) Representative images of WT and ΔVSP-1/2 strains with the vector or pDcdV. (**E**) Violin plots of cell length distributions of WT, ΔVSP-1/2, ΔVSP-1, and ΔVSP-2 strains with the vector or pDcdV: summary statistic for this and all following violin plots are mean (diamonds), median (horizontal black line), interquartile range (box), and data below and above the interquartile range (vertical lines). Different letters indicate significant differences (*n*=3) at $p < 0.05$, according to Tukey's post-hoc test.

Fig. 2

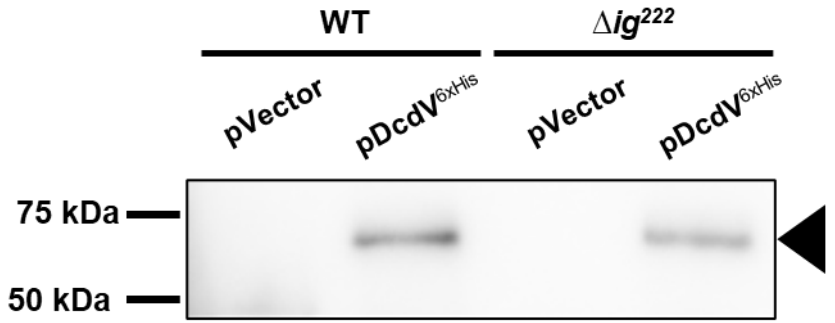
A



B



C



D

difV constructs size (position in ig^{222})	Diagram of *difV* constructs (5' → 3')	Inhibits DcdV?
222 nt (ig^{222})	RBS 1 111 222	YES
222 nt^{STOP}	*	YES
174 nt (49-222)		YES
174 nt^{-RBS}		YES
174 nt^{STOP}	*	YES
174 nt^{InteriorSTOP}	*	YES
138 nt (49-186)		NO
156 nt (49-204)		NO
166 nt (49-214)		NO
170 nt (49-218)		NO
157 nt (66-222)		NO
137 nt (86-222)		NO
100 nt (123-222)		NO
npcR_3991 (102-206)		NO

Fig. 2: DifV is a sRNA that post-translationally regulates DcdV

(**A**) Distribution of cell lengths measured from three biological replicates of gene deletions within VSP-1 or (**B**) individual gene deletions as indicated containing vector or pDcdV grown in the presence of 100 µM IPTG for 8 h. Different letters indicate significant differences (n=3) at $p < 0.05$, according to Tukey's post-hoc test. (**C**) Representative anti-6x His antibody Western blot of whole cell lysates from *V. cholerae* WT and Δ*ig*^{222} cultures maintaining vector or pDcdV^{6xHis}. Analysis was performed in triplicate biological samples. Black triangle corresponds to DcdV^{6xHis} (60.6 kDa). (**D**) Table of various *difV* constructs expressed in Δ*ig*^{222} under a P_{tac}-inducible promoter with a non-native ribosomal binding site (RBS, denoted by dotted line). DcdV induced filamentation in the presence of these *difV* constructs was assessed using fluorescence microscopy in biological triplicate cultures. "*****" indicates a stop codon introduced in place of a putative start codon.

Fig. 3

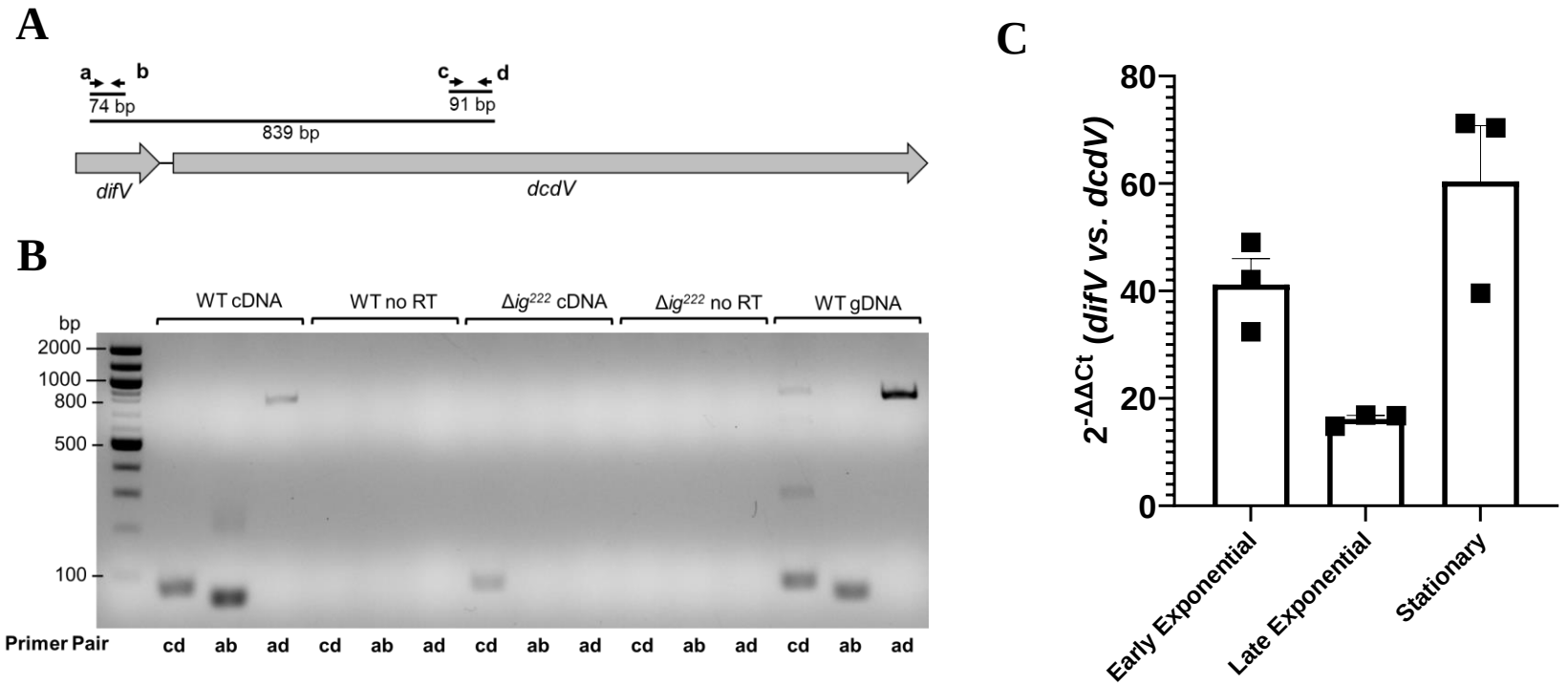

A
a
b
c
d
74 bp
91 bp
839 bp
difV
dcdV
B
WT cDNA
WT no RT
Δig222 cDNA
Δig222 no RT
WT gDNA
bp
2000
1000
800
500
100
Primer Pair
cd
ab
ad
C
2-ΔΔCt (difV vs. dcdV)
80
60
40
20
0
Early Exponential
Late Exponential
Stationary

Fig. 3: *difV* and *dcdV* are in an operon and *difV* expression exceeds *dcdV*

(**A**) Genomic diagram of *difV* and *dcdV* and the primers (a, b, c, and d) used for generating diagnostic PCR products. (**B**) PCR products amplified from nucleic acid templates (above) using the indicated primer pairs (below) resolved in a 1% agarose gel. All reactions were performed in duplicate using biologically independent samples with similar results. No RT = non-reverse transcribed RNA control. gDNA = genomic DNA control (**C**) qRT-PCR analysis of relative difference between *difV* transcript and *dcdV* transcript levels at different growth phases in WT *V. cholerae* normalized to an endogenous *gyrA* control. Data are graphed as mean ± SEM, *n*= 3.

Fig. 4

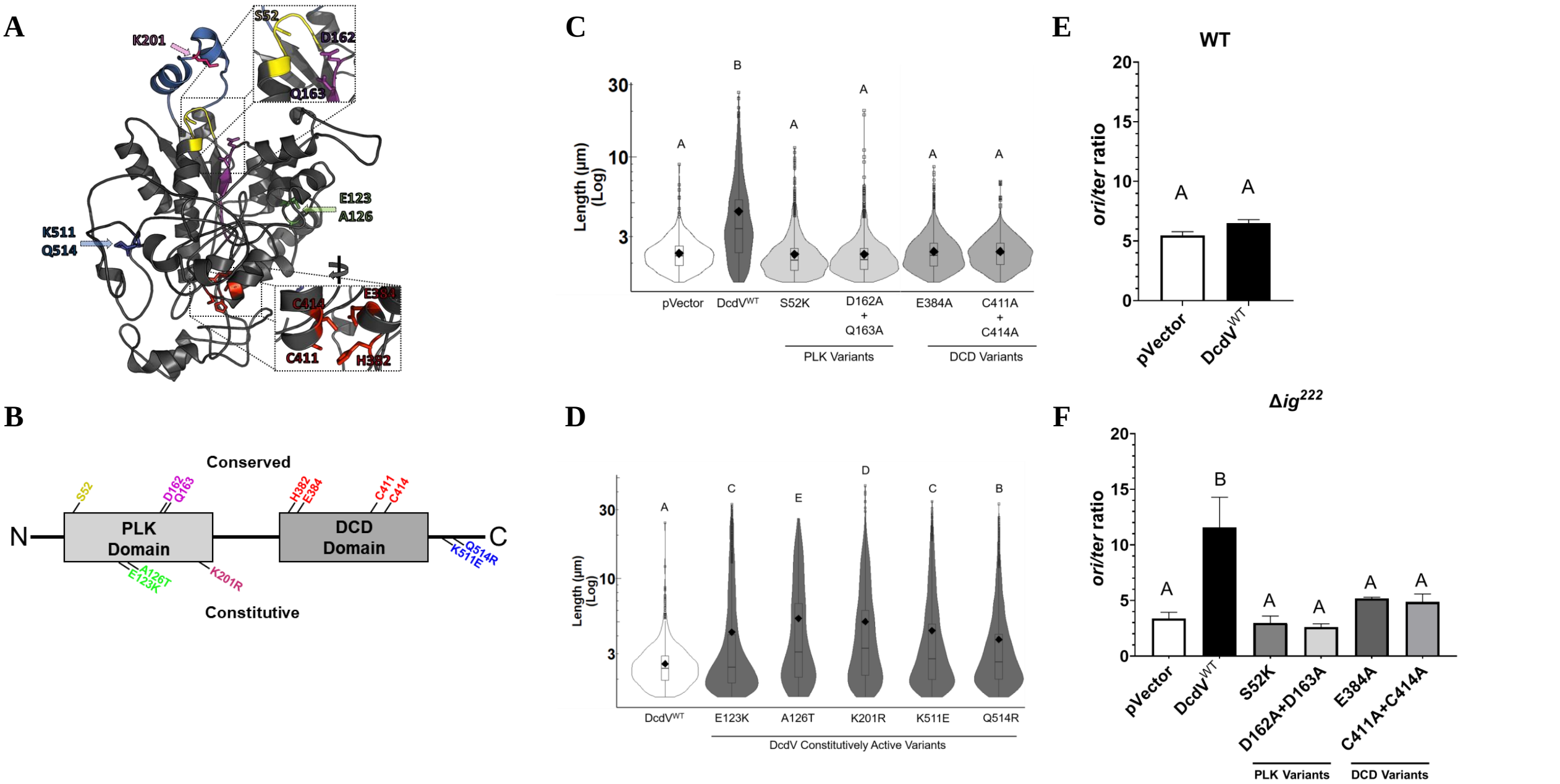
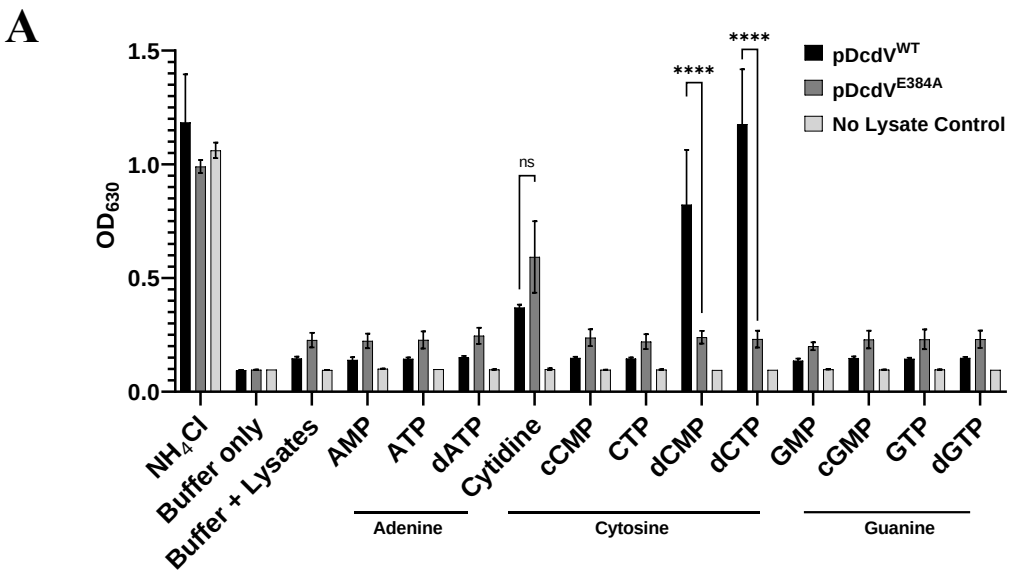
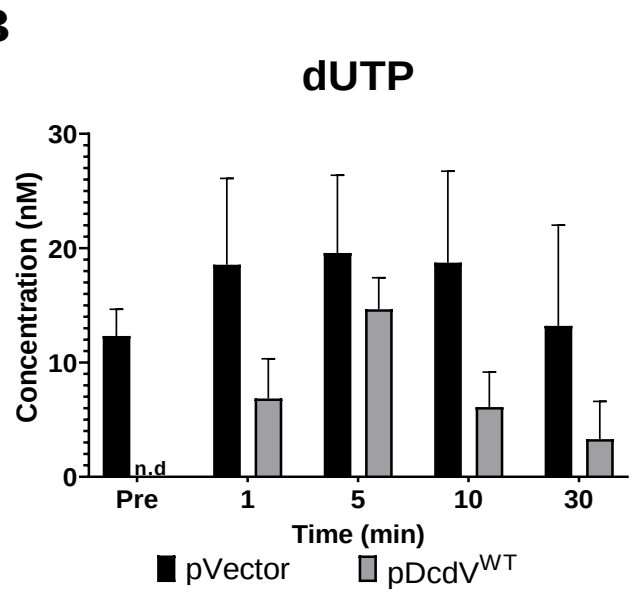

A
S52
D162
K201
Q163
E123
A126
K511
Q514
C414
E384
C411
H382
B
Conserved
S52
D162
Q163
H382
E384
C411
C414
N
PLK Domain
DCD Domain
C
Q514R
K511E
A126T
E123K
K201R
Constitutive
C
Length (µm) (Log)
30
10
3
A
B
A
A
A
A
pVector
DcdV^WT
S52K
D162A + Q163A
E384A
C411A + C414A
PLK Variants
DCD Variants
D
A
C
E
D
C
B
DcdV^WT
E123K
A126T
K201R
K511E
Q514R
DcdV Constitutively Active Variants
E
WT
ori/ter ratio
20
15
10
5
0
A
A
pVector
DcdV^WT
F
Δig^222
B
A
A
A
A
A
pVector
DcdV^WT
S52K
D162A+D163A
E384A
C411A+C414A
PLK Variants
DCD Variants

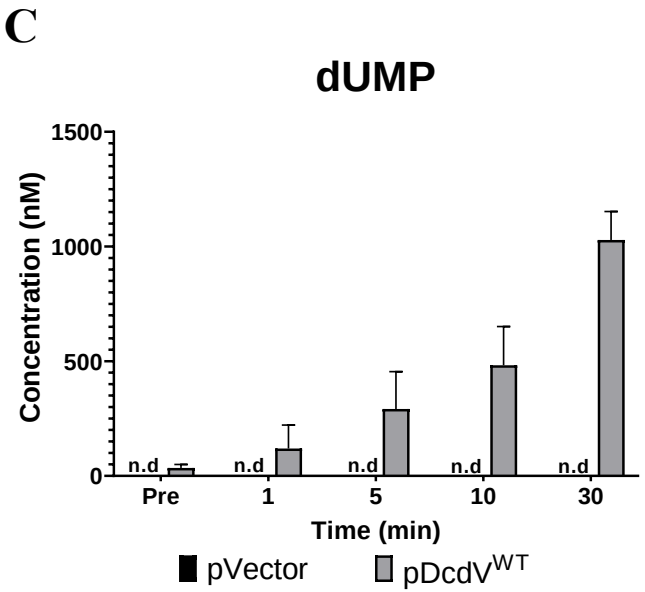
Fig. 4: Both the PLK and DCD domains are required for DcdV induced filamentation

(**A**) Phyre2 predicted structure of DcdV from *V. cholerae* El Tor. The inset shows the conserved residues of PLK (top) and DCD (bottom) domains. (**B**) Domain organization and conserved residues at each domain of DcdV. Top labeled residues indicate conserved features of both domains, and the bottom labeled residues indicate variants that render DcdV constitutively active. (**C**) Distribution of cell lengths measured from three biological replicates of WT *E. coli* as indicated. (**D**) Distribution of cell lengths measured from three biological replicates of the Δ*dcdV V. cholerae* mutant expressing the indicated DcdV variants. *ori/ter* ratios of Chromosome 1 in (**E**) WT and (**F**) Δig^{222} *V. cholerae* strains expressing the indicated DcdV construct for 8 h and quantified using qRT-PCR. Each bar represents the mean ± SEM, *n*=3. Different letters indicate significant differences (*n*=3) at $p < 0.05$, according to Tukey's post-hoc test.

Fig. 5


A
1.5
1.0
0.5
0.0
OD630
pDcdV^WT
pDcdV^E384A
No Lysate Control
****
****
ns
NH4Cl
Buffer only
Buffer + Lysates
AMP
ATP
dATP
Cytidine
cCMP
CTP
dCMP
dCTP
GMP
cGMP
GTP
dGTP
Adenine
Cytosine
Guanine


B
dUTP
Concentration (nM)
30
20
10
0
n.d
Pre
1
5
10
30
Time (min)
pVector
pDcdV^WT


C
dUMP
Concentration (nM)
1500
1000
500
0
n.d
Pre
1
5
10
30
Time (min)
pVector
pDcdV^WT

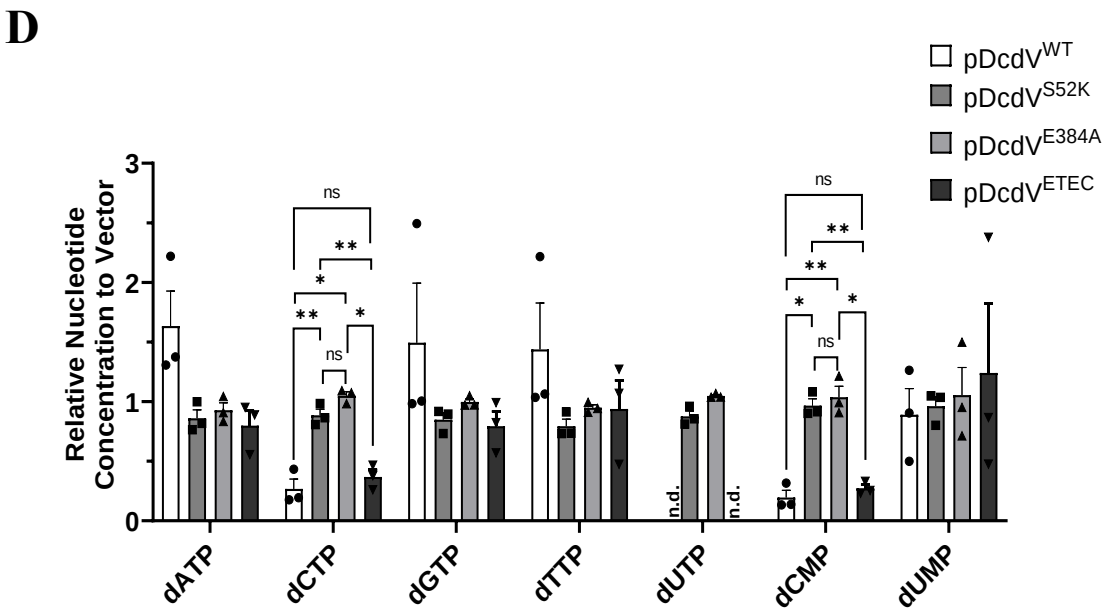

D
pDcdV^WT
pDcdV^S52K
pDcdV^E384A
pDcdV^ETEC
Relative Nucleotide Concentration to Vector
3
2
1
0
ns
**
*
**
ns
*
n.d.
n.d.
ns
**
*
**
ns
*
dATP
dCTP
dGTP
dTTP
dUTP
dCMP
dUMP

Fig. 5: DcdV alters cellular nucleotide metabolism.

(**A**) Lysates collected from *E. coli* expressing DcdV or DcdV^{E384A} and a “no lysate” buffer control incubated with 12 nucleotide substrates (1.9 mM NH_4Cl as a positive control, 37.7 mM cytidine, and 7.5 mM for all other substrates). Data represent the mean ± SEM, *n*=3. Quantification of dUTP (**B**) and dUMP (**C**) using UPLC-MS/MS, in the indicated cell lysates before (Pre) and after addition of 1 mM dCTP. Each lysate was normalized to 20 mg/mL total protein. Each bar represents mean ± SEM, *n*=3. (**D**) Quantification of the indicated dNTPs in vivo using UPLC-MS/MS in strains expressing the four DcdV variants, as indicated, normalized to dNTP concentrations measured in a vector control. Data are graphed as mean ± SEM, *n*= 3, Two-way ANOVA with Tukey’s multiple-comparison test, normalized to pVector, n.d. indicates “none detected”, and ns indicates “not significant”.

Fig. 6

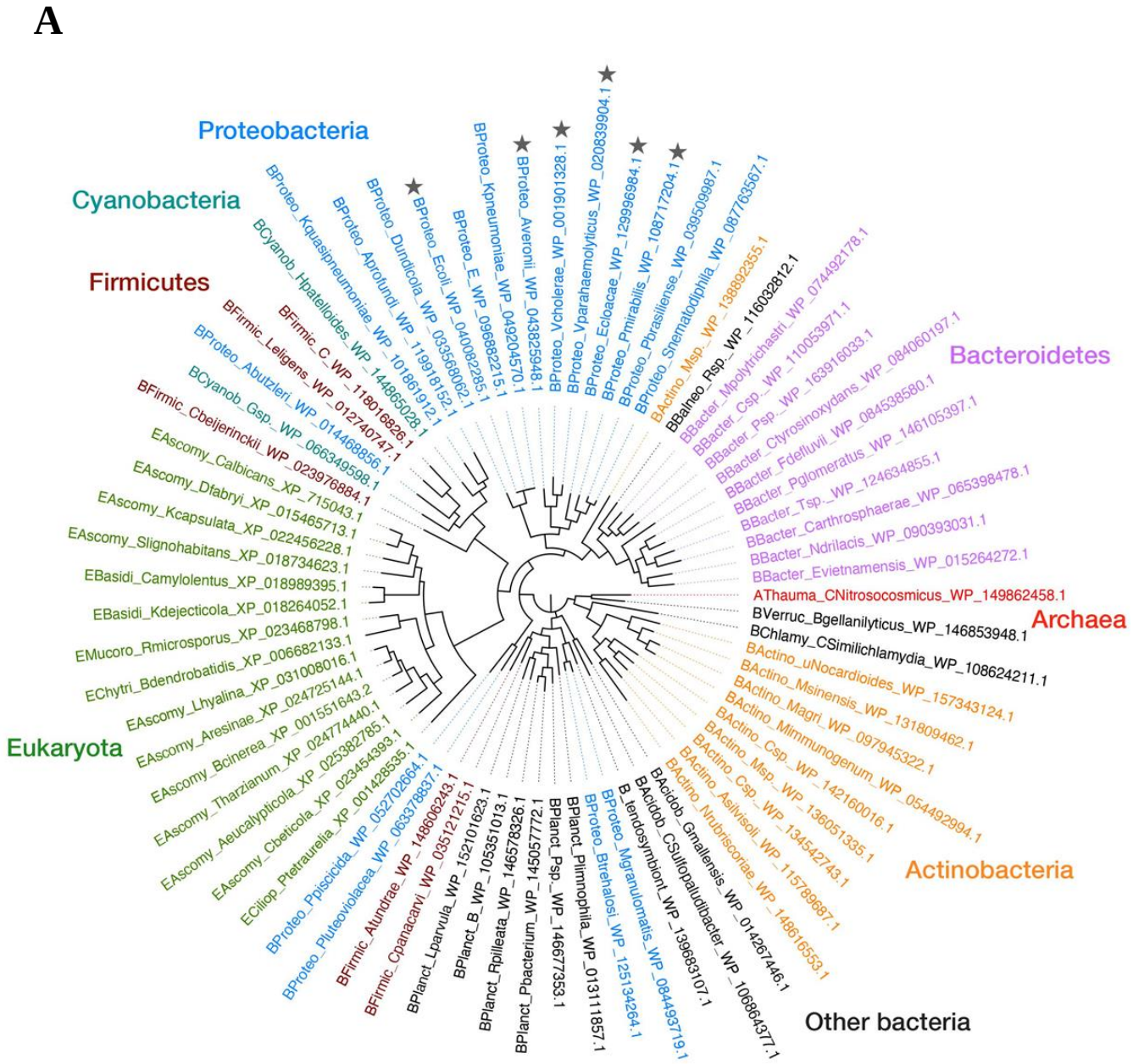

A
Proteobacteria
Cyanobacteria
Firmicutes
Bacteroidetes
Archaea
Eukaryota
Actinobacteria
Other bacteria
BProteo_Kquasipneumoniae_WP_101861912.1
BProteo_Aprofundi_WP_119918152.1
BProteo_Dundicola_WP_033568062.1
BProteo_Ecoli_WP_040082285.1
BProteo_E_WP_096882215.1
BProteo_Kpneumoniae_WP_049204570.1
BProteo_Averonii_WP_043825948.1
BProteo_Vcholerae_WP_001901328.1
BProteo_Vparahaemolyticus_WP_020839904.1
BProteo_Ecloacae_WP_129996984.1
BProteo_Pmirabilis_WP_108717204.1
BProteo_Pbrasiliense_WP_039509987.1
BProteo_Snematodiphila_WP_087763567.1
BCyanob_Hpatelloides_WP_144865028.1
BFirmic_C_WP_118016826.1
BFirmic_Leligens_WP_012740747.1
BProteo_Abutzleri_WP_014468856.1
BCyanob_Gsp._WP_066349598.1
BFirmic_Cbeijerinckii_WP_023976884.1
EAscomy_Calbicans_XP_715043.1
EAscomy_Dfabryi_XP_015465713.1
EAscomy_Kcapsulata_XP_022456228.1
EAscomy_Slignohabitans_XP_018734623.1
EBasidi_Camylolentus_XP_018989395.1
EBasidi_Kdejecticola_XP_018264052.1
EMucoro_Rmicrosporus_XP_023468798.1
EChytri_Bdendrobatidis_XP_006682133.1
EAscomy_Lhyalina_XP_031008016.1
EAscomy_Aresinae_XP_024725144.1
EAscomy_Bcinerea_XP_001551643.2
EAscomy_Tharzianum_XP_024774440.1
EAscomy_Aeucalypticola_XP_025382785.1
EAscomy_Cbeticola_XP_023454393.1
ECiliop_Ptetraurelia_XP_001428535.1
BProteo_Ppiscicida_WP_052702664.1
BProteo_Plutecoviolacea_WP_063378837.1
BFirmic_Atundrae_WP_148606243.1
BFirmic_Cpanacanvi_WP_035121215.1
BPlanct_Lparvula_WP_152101623.1
BPlanct_B_WP_105351013.1
BPlanct_Rpilleata_WP_146578326.1
BPlanct_Pbacterium_WP_145057772.1
BPlanct_Psp._WP_146677353.1
BPlanct_Plimnophila_WP_013111857.1
BProteo_Btrehalosi_WP_125134264.1
BProteo_Mgranulomatis_WP_084493719.1
B_tendosymbiont_WP_139683107.1
BAcidob_CSulfopaludibacter_WP_106864377.1
BAcidob_Gmallensis_WP_014267446.1
BActino_Nrubriscoriae_WP_148616553.1
BActino_Asilvisoli_WP_115789687.1
BActino_Csp._WP_134542743.1
BActino_Msp._WP_136051335.1
BActino_Csp._WP_142160016.1
BActino_Mimmunogenum_WP_054492994.1
BActino_Magri_WP_097945322.1
BActino_Msinensis_WP_131809462.1
BActino_uNocardioides_WP_157343124.1
BChlamy_CSimilichlamydia_WP_108624211.1
BVerruc_Bgellanilyticus_WP_146853948.1
AThauma_CNitrosocosmicus_WP_149862458.1
BBacter_Evietnamensis_WP_015264272.1
BBacter_Ndrilacis_WP_090393031.1
BBacter_Carthrosphaerae_WP_065398478.1
BBacter_Tsp._WP_124634855.1
BBacter_Pglomeratus_WP_146105397.1
BBacter_Fdefluvii_WP_084538580.1
BBacter_Ctyrosinoxydans_WP_084060197.1
BBacter_Psp._WP_163916033.1
BBacter_Csp._WP_110053971.1
BBacter_Mpolytrichastri_WP_074492178.1
BActino_Msp._WP_138892355.1
BBalneo_Rsp._WP_116032812.1
B
Length (μm) (Log)
30
10
3
A
C
A
C
A
E
B
D
A
DcdV Vector VC VC VP VP PM PM ETEC ETEC
DifV Vector Vector VC Vector VP Vector PM Vector ETEC
C
DcdV
DifV
VC
VP
PM
ETEC
VC
VP
PM
ETEC
2 μm

Fig. 6: *dcdV* and *difV* are widely conserved

(**A**) Phylogenetic tree of DcdV homologs containing PLK and DCD domains from representative phyla across the tree of life. Stars indicate query proteins of interest in this study. (**B**) Distribution of cell lengths expressing the indicated DcdV homologs and their cognate DifV or vector control in *E. coli* (n=3). Different letters indicate significant differences (n=3) at p< 0.05, according to Dunnett's post-hoc test against the control (pVector^{DcdV} + pVector^{DifV}) strain. (**C**) Representative images of *E. coli* expressing pDcdV/homologs and pDifV/homologs combinations. Scale represents 2 µm.

Fig. 7

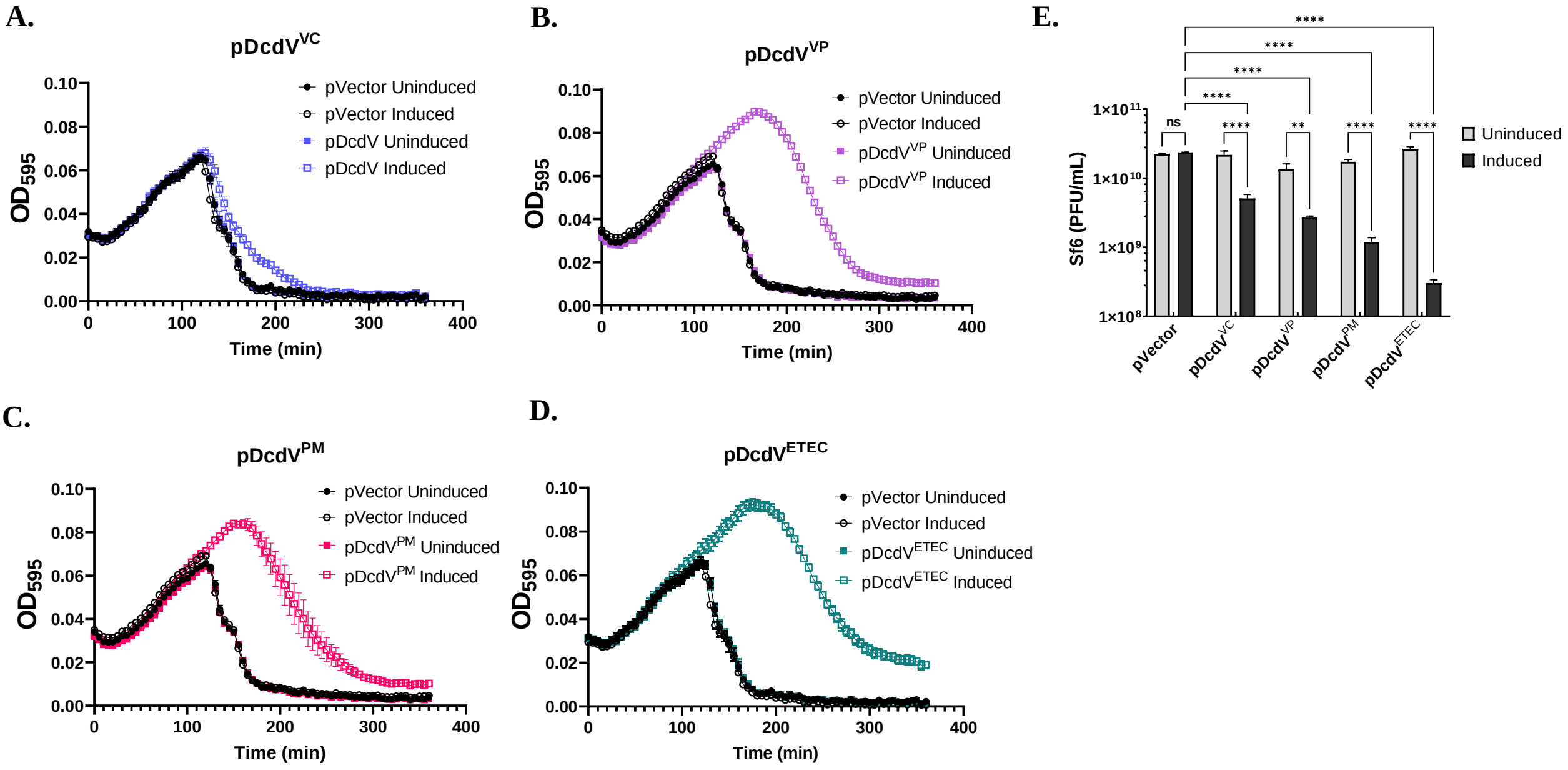
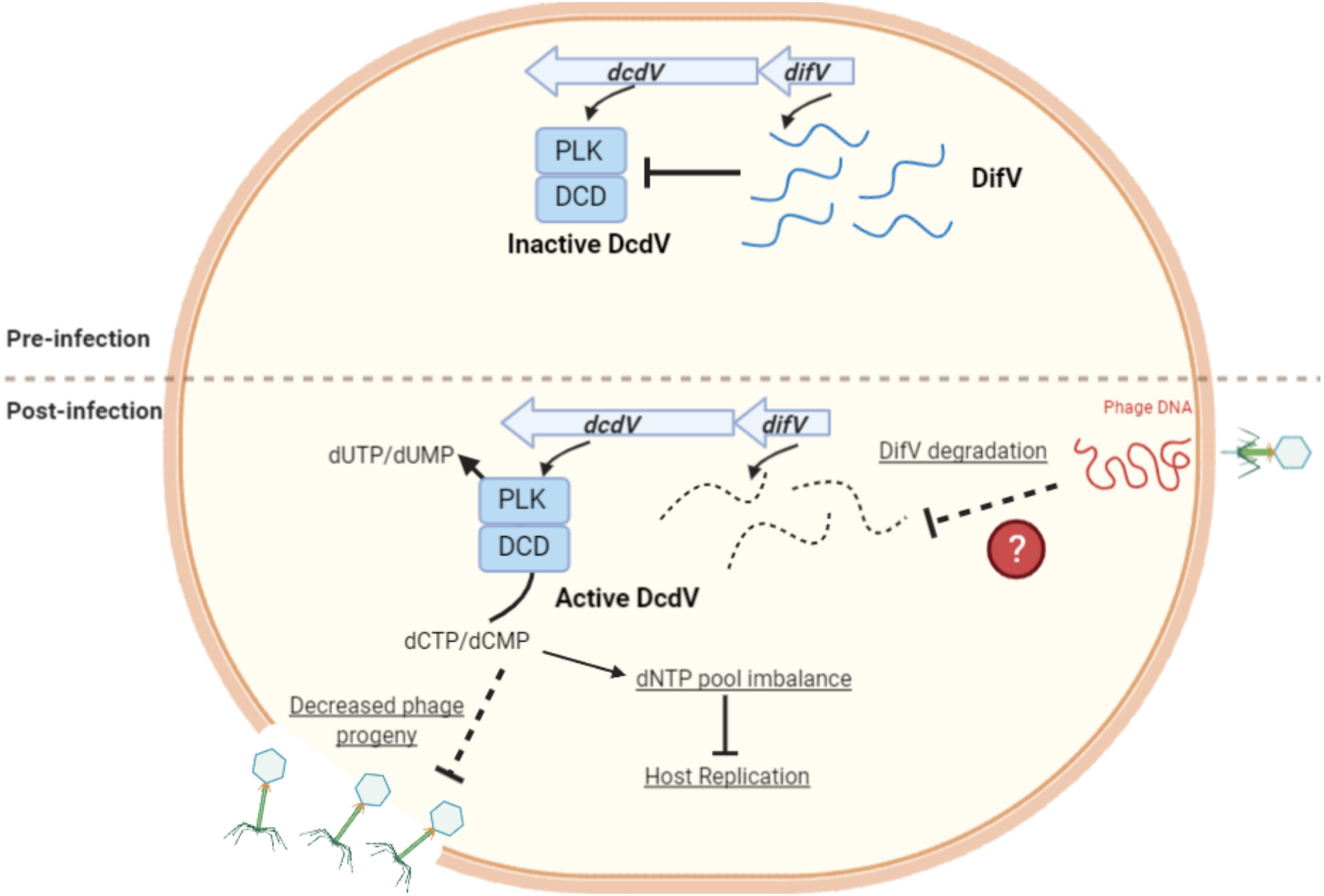

A.
pDcdV^VC
pVector Uninduced
pVector Induced
pDcdV Uninduced
pDcdV Induced
OD595
Time (min)
B.
pDcdV^VP
pVector Uninduced
pVector Induced
pDcdV^VP Uninduced
pDcdV^VP Induced
OD595
Time (min)
C.
pDcdV^PM
pVector Uninduced
pVector Induced
pDcdV^PM Uninduced
pDcdV^PM Induced
OD595
Time (min)
D.
pDcdV^ETEC
pVector Uninduced
pVector Induced
pDcdV^ETEC Uninduced
pDcdV^ETEC Induced
OD595
Time (min)
E.
Sf6 (PFU/mL)
ns
****
**
Uninduced
Induced
pVector
pDcdV^VC
pDcdV^VP
pDcdV^PM
pDcdV^ETEC

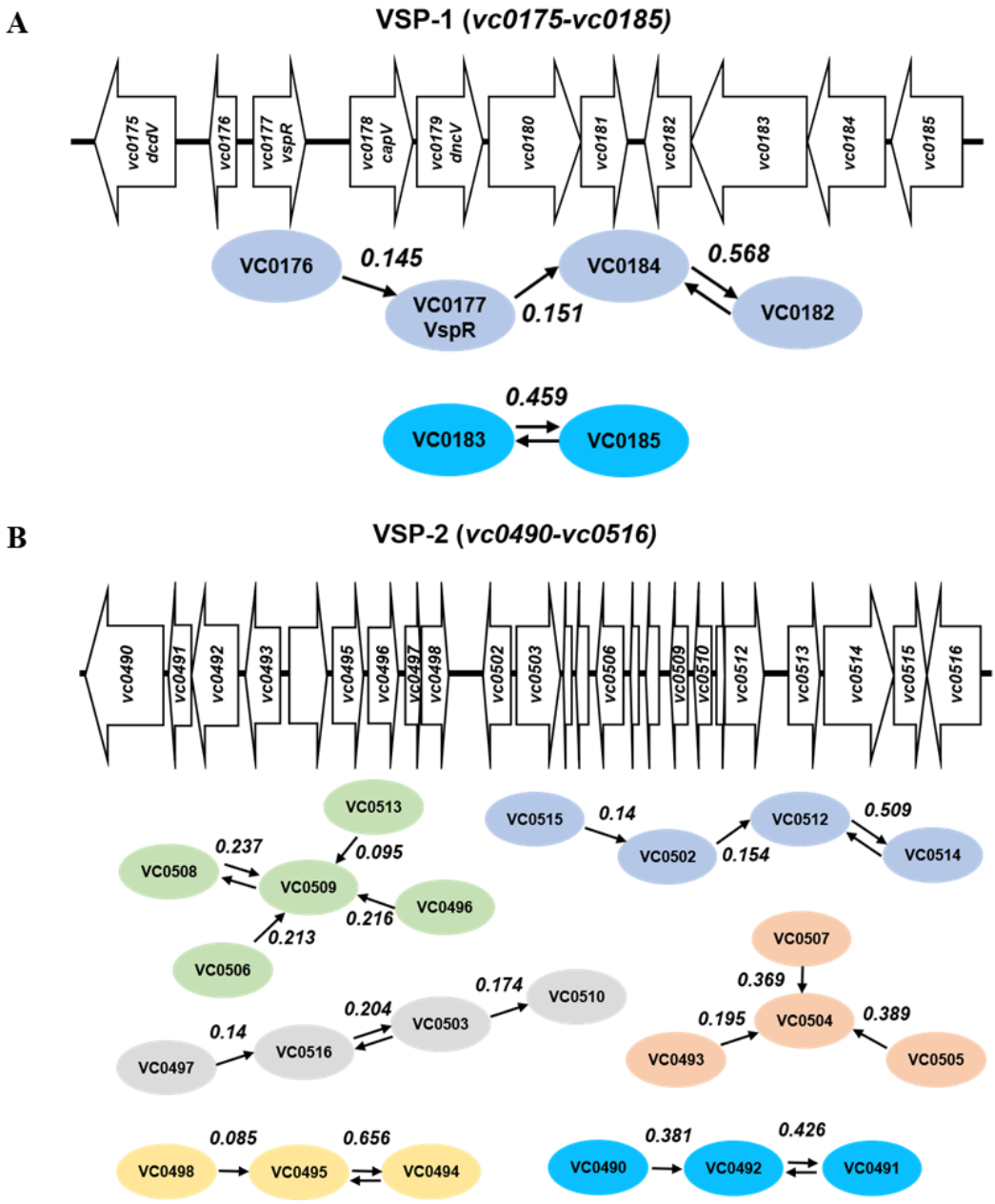
Fig. 7: DcdV mediates phage defense

(**A-D)** Growth curves for *S. flexneri* containing vector or pDcdV/homologs infected Sf6 at time 0 at an MOI of 0.1 in the presence or absence of 100 µM IPTG. Each graph represents three biological replicates each with three technical replicates. (**B**) Plaque-forming units (PFU) per mL of phage Sf6 measured at the conclusion of the *S. flexneri* growth curve experiment above. Results are represented as mean ± SEM, *n*= 3, Two-way ANOVA with Tukey's multiple-comparison test.

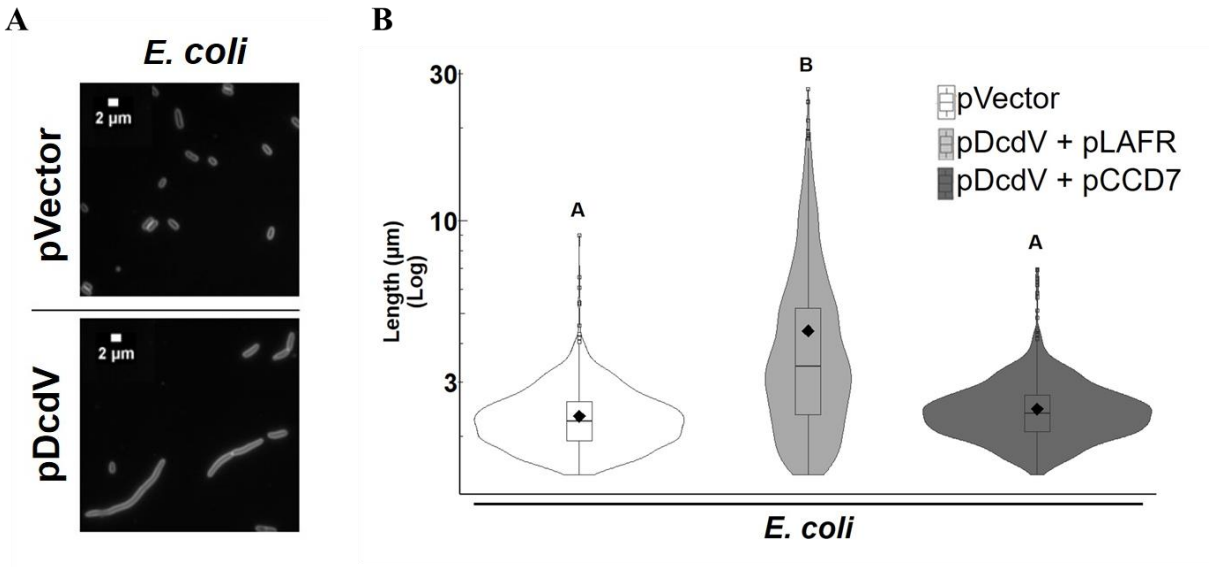
Graphical Abstract



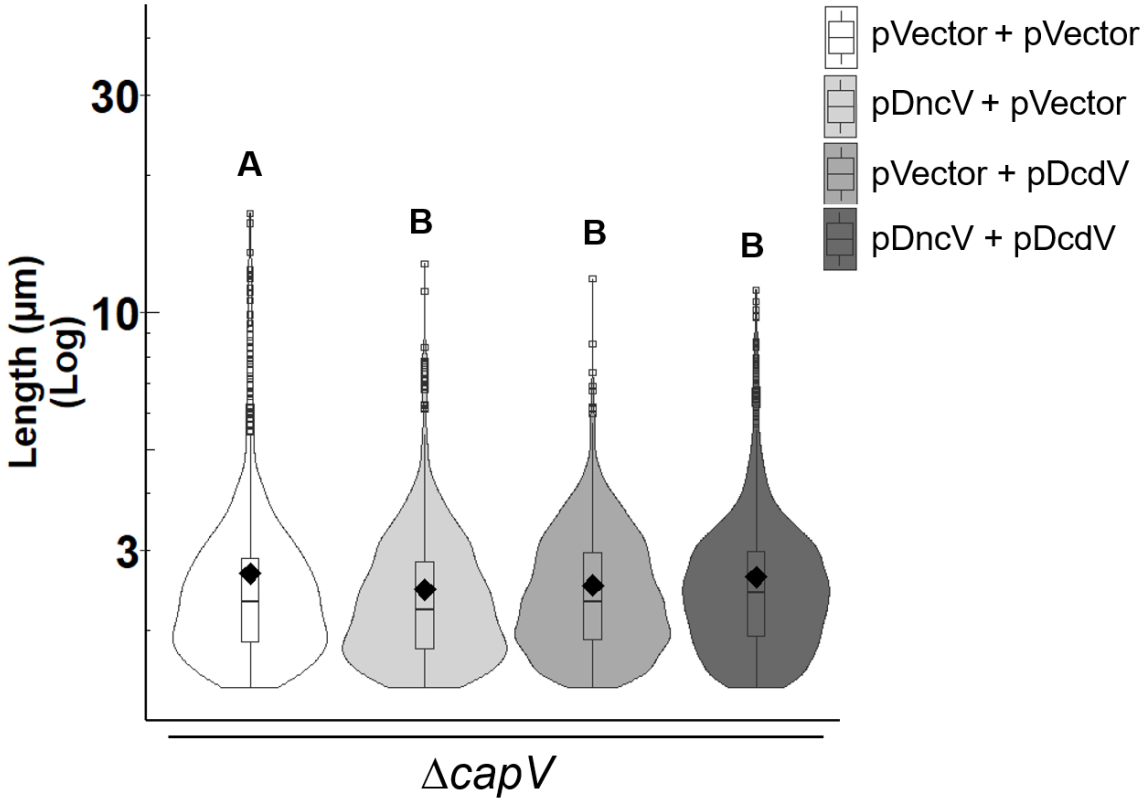
SUPPLEMENTAL MATERIAL



Supplemental Figure 1. VSP-1 and VSP-2 schematic and predicted maximum related subnetworks (MRS).
(**A**) Cartoon schematic and gene network predictions, other than DcdV and CBASS (see Figs. 1B and 1), of VSP-1 from El Tor *V. cholerae* N16961 (not to scale). (**B**) Cartoon schematic and gene network predictions of VSP-2 from El Tor *V. cholerae* N16961 (not to scale). Arrows indicate the highest partial correlation W_{ij} of each individual VSP gene to another (represented by ovals). Two arrows pointing in opposing directions indicates the two genes each have their highest correlation to each other.

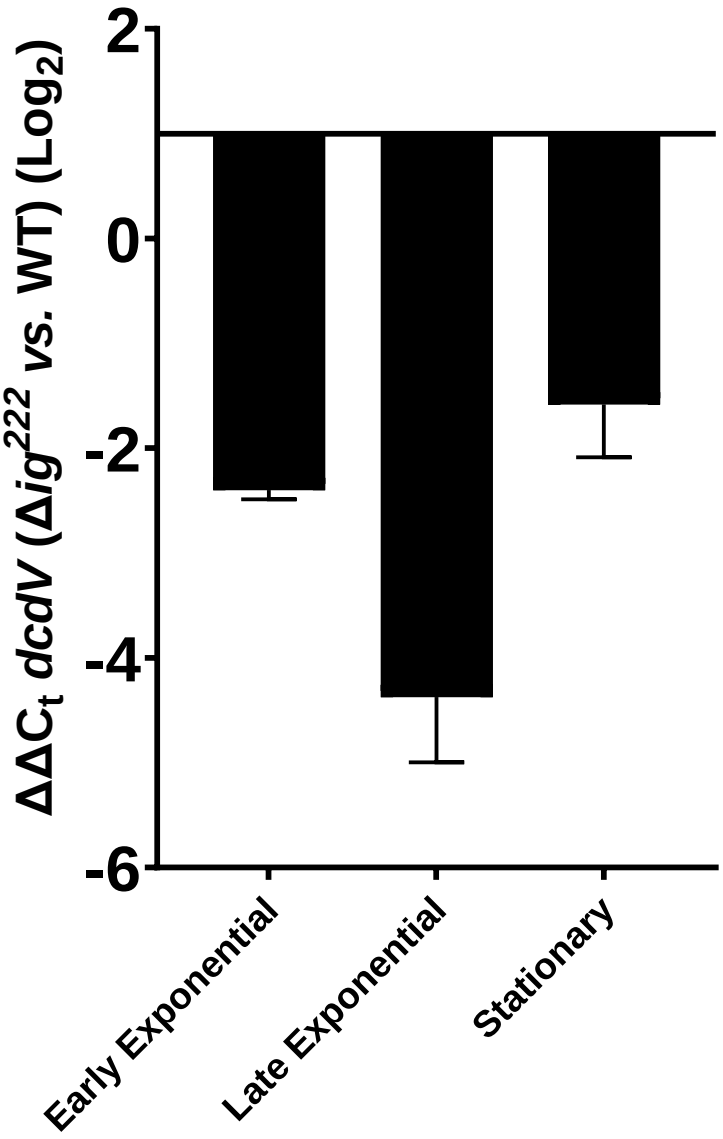


Supplemental Figure 2. Ectopic expression of *dcdV* leads to cell filamentation in *E. coli* that is alleviated by provision of a single copy cosmid containing VSP-1.
(**A**) Representative images of *E. coli* cultures maintaining an empty vector plasmid (pVector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 µM IPTG for 8 h. Cells were stained with FM4-64 prior to imaging. Scale represents 2 µm. (**B**) Distribution of cell lengths measured from three biological replicates of *E. coli* cultures carrying an empty vector (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) in addition to either an empty vector single copy cosmid control (pLAFR) or pLAFR containing VSP-1 (pCCD7) grown in the presence of 100 µM IPTG for 8 h. Distributions represent ~1000 to 2000 cells measured per strain. Different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test.

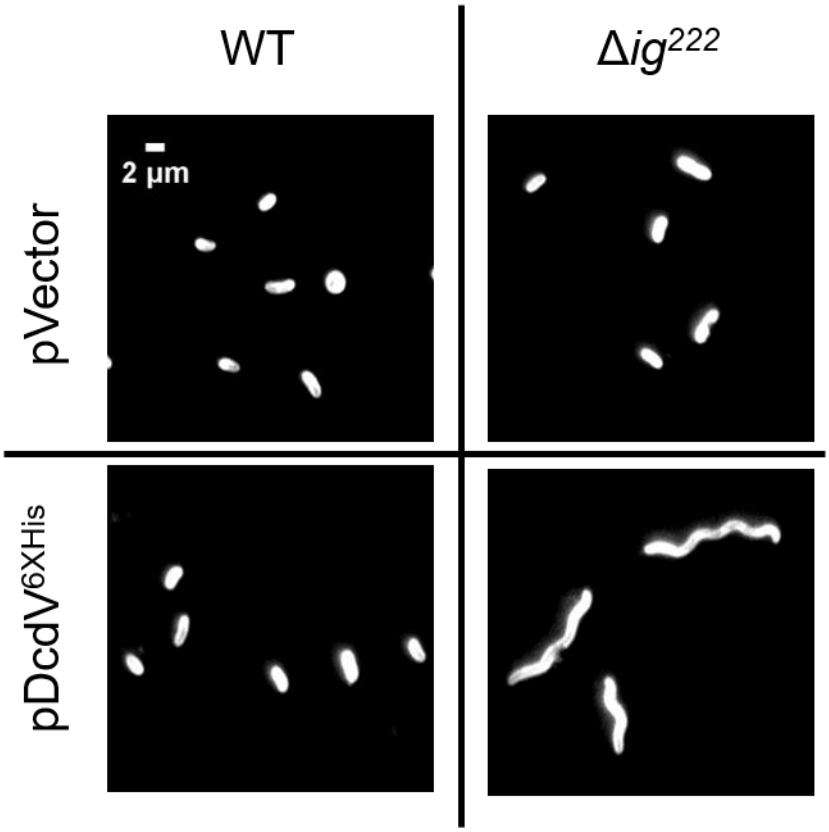


Supplemental Figure 3. Ectopic expression of DncV and DcdV does not lead to filamentation in the Δ*capV* mutant of *V. cholerae.*

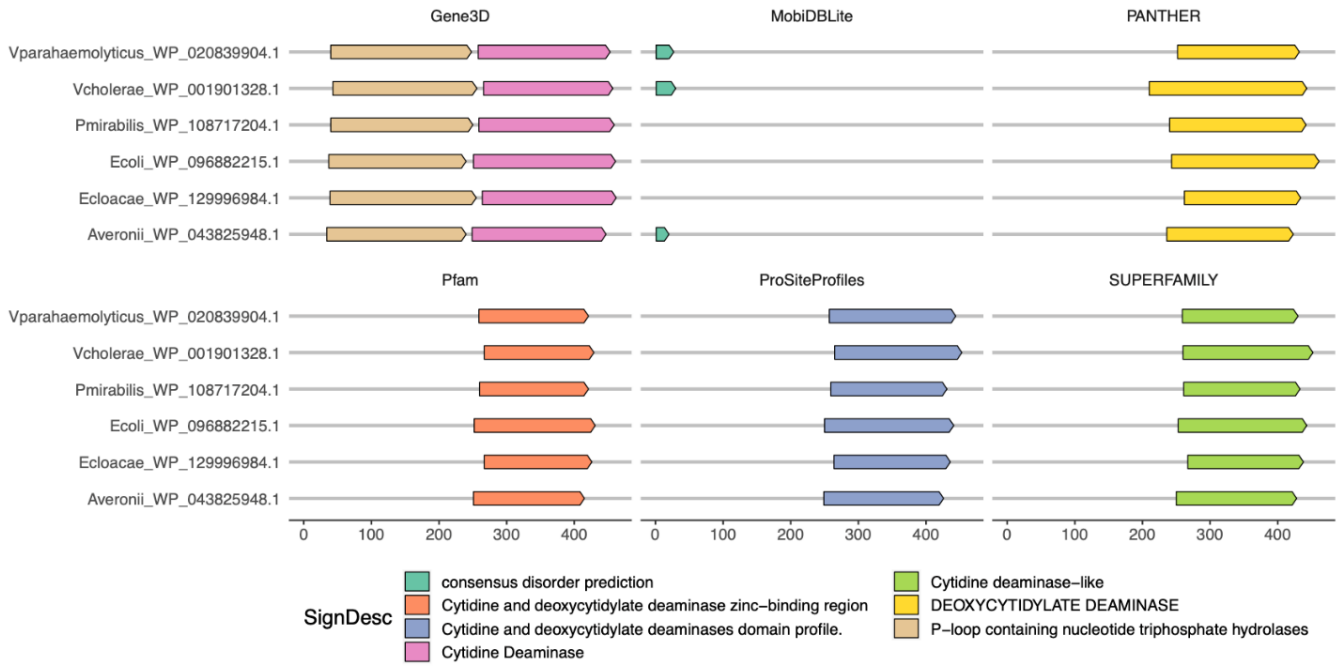
Distribution of cell lengths measured from three biological replicates of *ΔcapV* mutant cultures maintaining the indicated plasmids grown in the presence of 100 µM IPTG for 8 h. Distributions represent ~1200-1700 cells measured per strain. Different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test.



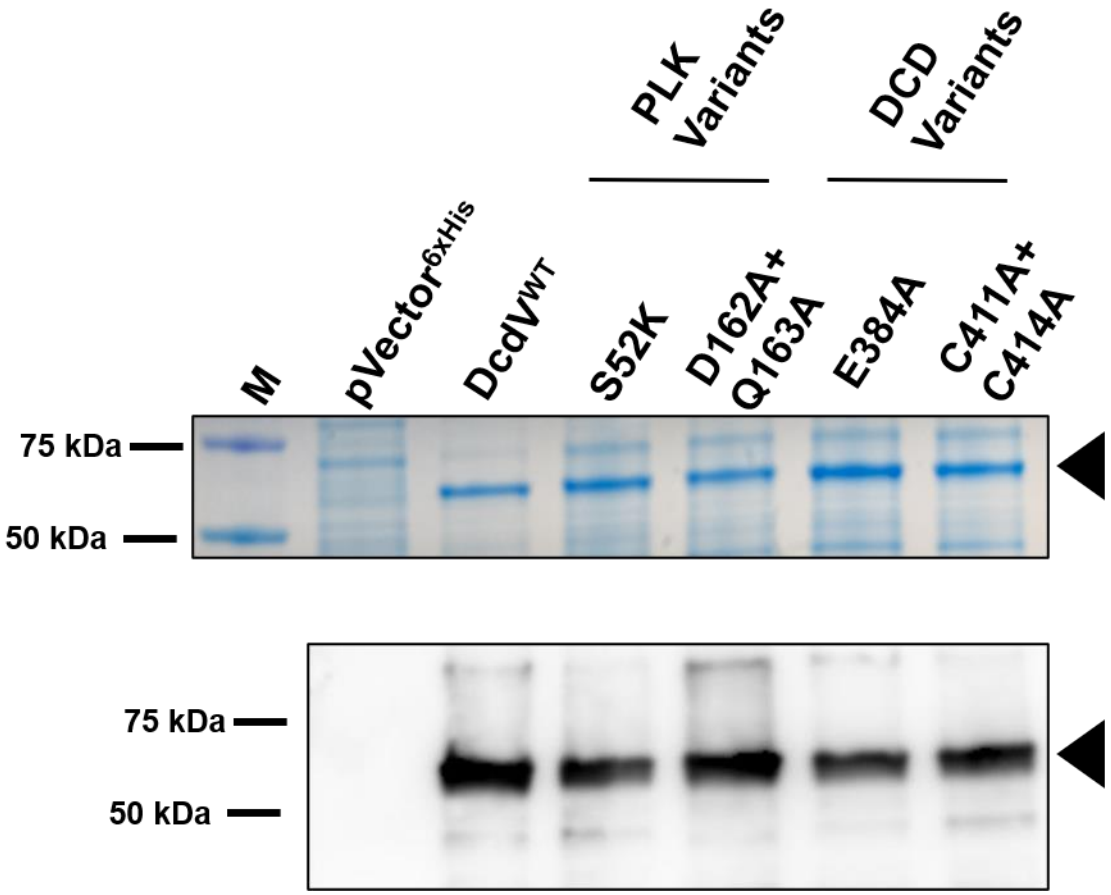
Supplemental Figure 4. Δ*ig*^{222} has decreased *dcdV* expression relative to WT *V. cholerae*. Relative difference in *dcdV* expression between Δ*ig*^{222} and WT *V. cholerae* at three different growth phases using qRT-PCR and an endogenous *gyrA* control. Data represent the mean ± SEM of three biological replicates.



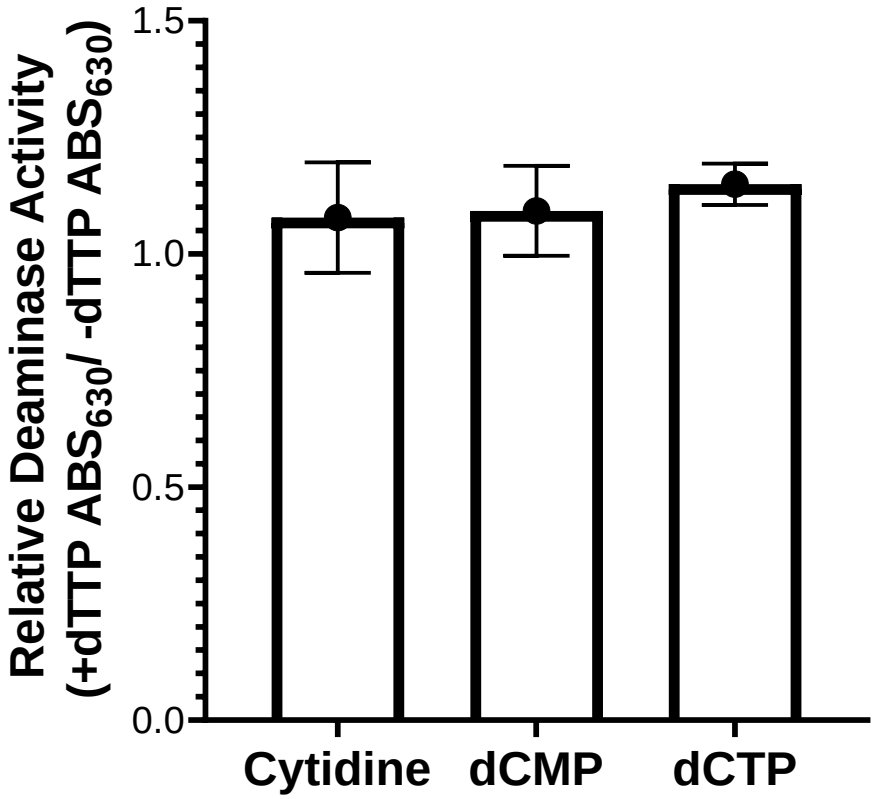
Supplemental Figure 5. DcdV C-terminal 6x Histidine fusion maintains the same activity as the WT DcdV enzyme.
Representative images of *V. cholerae* WT and Δ*ig*^{222} cultures maintaining an empty vector plasmid (pVector) or P_{tac}-inducible *dcdV-6xHIS* plasmid (pDcdV^{6xHis}) grown in the presence of 100 µM IPTG for 2 h. Cells were stained with FM4-64 prior to imaging and performed in biological triplicate.



Supplemental Figure 6. Domain architectures of the six DcdV query proteins.
Domain architecture and secondary structure predictions for the six proteobacterial starting points of interest (query proteins) using InterProScan [[40]; see Methods]. Results from six main analyses are shown here for the query proteins: Gene3D (including CATH structure database), Pfam, ProSiteProfiles, PANTHER, and SUPERFAMILY protein domain profile databases, and MobiDBLite for disorder prediction. No transmembrane regions (using TMHMM) or membrane/extracellular localization were predicted for any of the proteins (using Phobius); hence not shown.



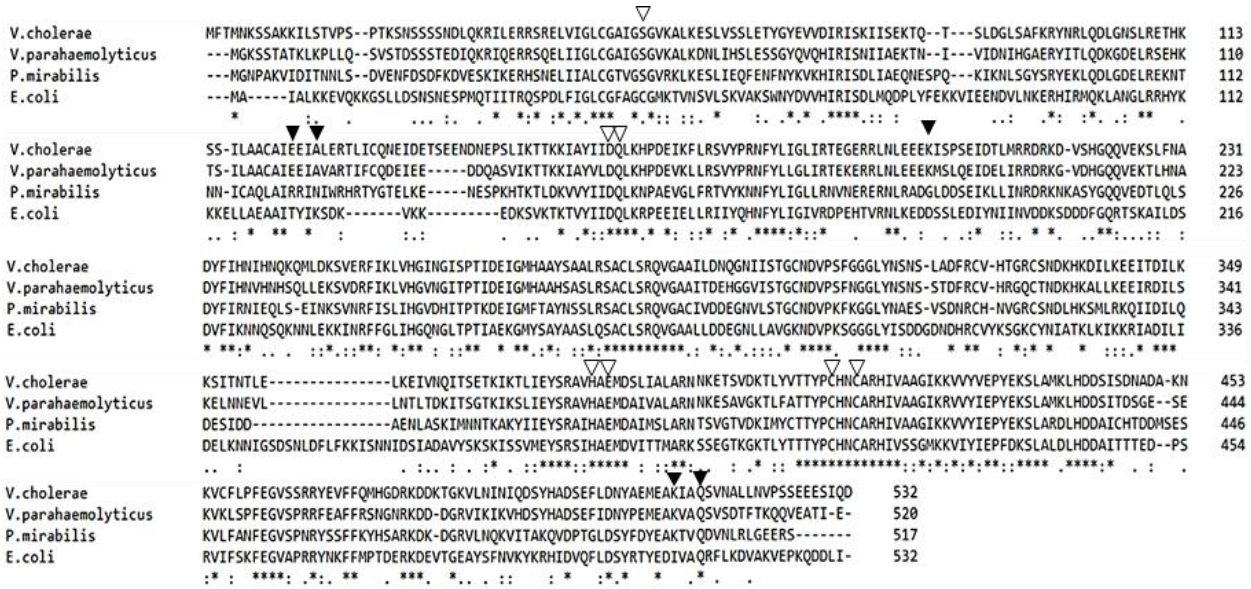
Supplemental Figure 7. Cellular abundance of C-terminal 6x histidine tagged DcdV variant fusions analyzed by Coomassie stain and Western blot.
Representative Coomassie stained gel (top) and anti-6x His antibody Western blot (bottom) of whole cell lysates from *E. coli* BL21(DE3) cells maintaining an empty vector (pVector^{6xHis}), inducible C-terminal 6x histidine tagged *dcdV* (WT) or *dcdV* variants (S52K, D162A + Q163A, E384A, and C411A + C414A) grown in the presence of 1 mM IPTG for 3 h. Sample inputs were normalized by culture OD_{600} and resolved by SDS-PAGE. Three biological replicates of each strain were analyzed with similar results. Black triangles correspond to the predicted molecular weight of the DcdV tagged fusions (60.6 kDa). M = molecular weight marker.



Supplemental Figure 8. Addition of exogenous dTTP does not inhibit DcdV deaminase activity in *E. coli* lysates.
Lysates collected from *E. coli* expressing WT DcdV incubated with or without exogenous 7.5 mM dTTP and either 75 mM cytidine, 7.5 mM dCMP, or 7.5 mM dCTP. The evolution of NH_4^+ resulting from substrate deamination was detected by measuring the solution ABS_{630} after a Berthelot's reaction in microtiter plates. The relative deaminase activity was calculated by dividing the ABS_{630} of the +dTTP reaction by the no dTTP control reaction for each lysate. Data represent the mean ± SEM of three biological replicate lysates.

	Absolute Intracellular dNTP Concentration in pmol/mg of Total Protein				
Nucleotides	pVector	pDcdV^{WT}	pDcdV^{S52K}	pDcdV^{E384A}	pDcdV^{ETEC}
dATP	75.6 ± 5.3	121.9 ± 18.1	64.6 ± 3.6	70.1 ± 5.7	59.1 ± 5.8
dCTP	6.6 ± 0.7	1.6 ± 0.3	5.8 ± 0.3	6.9 ± 0.7	2.4 ± 0.2
dGTP	42.7 ± 4.5	59.9 ± 13.9	35.7 ± 1.8	42.4 ± 3.7	32.8 ± 1.9
dTTP	11.4 ± 0.7	16.5 ± 4.7	9.0 ± 0.3	10.80 ± 0.4	10.4 ± 2.2
dUTP	1.1 ± 0.1	n.d.	0.9 ± 0.1	1.1 ± 0.10	n.d.
dCMP	7158.3 ± 2485.5	1185.7 ± 227.9	6859.4 ± 2226.4	7027.8 ± 1950.4	1945.1 ± 636.2
dUMP	9.3 ± 2.90	7.1 ± 0.3	9.1 ± 3.3	8.7 ± 1.1	9.9 ± 3.3

Supplemental Figure 9. Absolute intracellular concentration of deoxynucleotides. Quantification of the indicated dNTPs in vivo, using UPLC-MS/MS, in strains expressing the empty vector and four DcdV variants, as indicated. Data represents mean ± SEM, *n*=3.



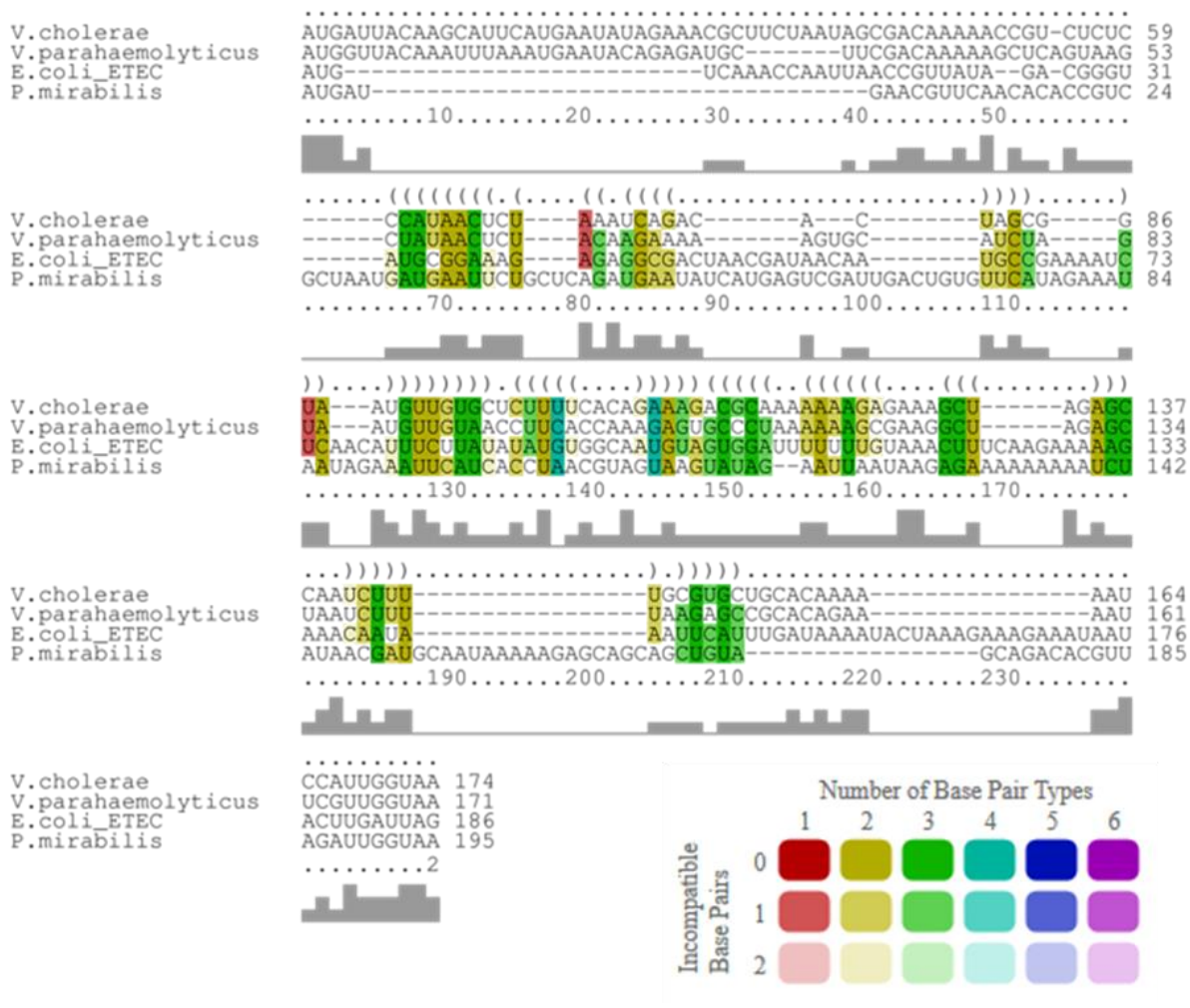
Supplemental Figure 10. ClustalW multiple sequence alignment of DcdV homologs explored in this study.

Amino acid alignment of DcdV and three homologs using webservice EMBL-EBI [74]. "*****" indicates 100% identity, "**:**" indicates >75%, and "**.**" Indicates >50% similarity. Open triangles above the alignments indicate conserved residues of PLK and DCD domains. Closed triangles indicate amino acids where single amino acid substitutions were found to render *V. cholerae* DcdV insensitive to DifV inhibition (Figs. 6A, B, and D).

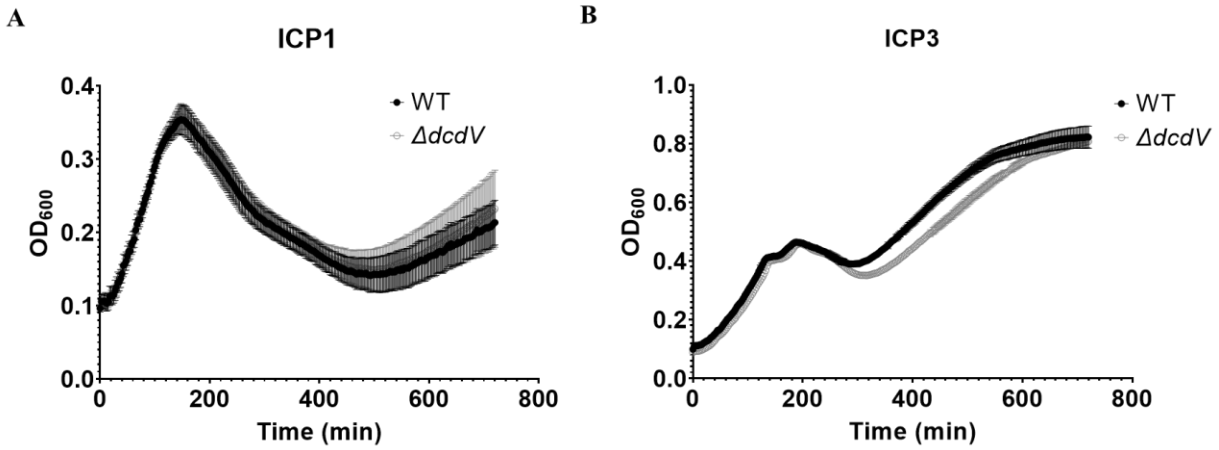
```
V.cholerae              ---MITSIH-----EYRNASNSDKNRLSHNSKSDTSGNVVLFSQKDAKKE--KARANLLR    50
V.parahaemolyticus      ---MVTNLN-----EYRDASTKSSVSYNS-TRKSASSNVVTFTKECPKKA--KARANLLR    49
P.mirabilis             ---MMNVQHTVANDEFCSDEYHESIDCVHRNNRNSSPNVVSIELIREKKIYNDAIKRAAA    57
E.coli_ETEC             MSNQLTVIDGYAEREATNDNNAE-----NLNISYICGNVVDFFVNFQEKETINSFDKILK    55
                            :.  .     *  . .  .       .    . *** :     :*   .:  .

V.cholerae              AAQKIHW 57
V.parahaemolyticus      AAQKIRW 56
P.mirabilis             VADTLDW 64
E.coli_ETEC             KEIILD- 61
                            :
```

Supplemental Figure 11. DifV (174 nt) and the three ORFs encoded upstream of *dcdV* homologs do not exhibit amino acid similarity.

Amino acid alignment of the *V. cholerae* Ig[222] translated ORF and three ORFs 5' of the *dcdV* homologs using EMBL-EBI ClustalW [74]. "*****" indicates 100% identity, "**:**" indicates >75%, and "**.**" Indicates >50% similarity.



Supplemental Figure 12. DifV (174 nt) and the three ORFs encoded upstream of *dcdV* homologs do not have exhibit similarity.
Nucleotide alignment of the *V. cholerae* DifV and the ORFs 5' of *dcdV* homologs using LocARNA [75]. Consensus identities are correlated with the height of the bars below the corresponding nucleotide (bottom). The average secondary structure is indicated in dot-bracket notation (top). Compatible base pairs are colored according to the number of different types C-G (1), G-C (2), A-U (3), U-A (4), G-U (5) or U-G (6) of compatible base pairs in the corresponding columns. The saturation decreases with the number of incompatible base pairs.



Supplemental Figure 13. *V. cholerae* lacking *dcdV* do not exhibit enhanced susceptibility to predation by *V. cholerae* lytic phage ICP1 and ICP3.
Growth curves for *V. cholerae* WT and Δ*dcdV* infected by lytic phage ICP1 (**A**) and ICP3 (**B**). Bacteria were infected at time 0 at an MOI of 0.1 in microtiter plates. Data represent the mean ± SEM, *n*=3.

Supplementary Table 1. Strains and phages used in this study.

Strains	Name in this Study	Relevant Characteristics	Source or reference
E. coli			
DH10b		*F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139Δ(ara, leu)7697 galU galK λrpsL nupG*	ThermoFisher Scientific
BW29427		*RP4-2(TetSkan1360::FRT)*, *thrB1004*, *lacZ58*(M15), *ΔdapA1341::[erm pir$^+$]*, *rpsL*(strR), *thi-*, *hsdS-*, *pro-*	Lab Stock
BL21(DE3)		*F- ompT hsdSB(rB -mB +) gal dcm (DE3)*	Lab Stock
078:H11 H10407 (ETEC)	ETEC	Wild type	[106]
V. cholerae			
C6706str2	WT	Wild type O1 El Tor; Sm^R	[107]
E7946		Clinical isolate obtained in Bahrain in 1978; Sm^R	[108]
CR01	ΔVSP-1	O1 El Tor ΔVSP-1	This study
CR02	ΔVSP-2	O1 El Tor ΔVSP-2	This study
CR03	ΔVSP-1/2	O1 El Tor ΔVSP-1/2	This study
BYH206	Δ*ig*^{222}	O1 El Tor Δ*ig*^{222} *between vc0175-vc0176 position in N16961 chromosome I [177,230-177,008]*	This study
BYH207	Δ*vc0176*	O1 El Tor Δ*vc0176*	This study
BYH255	Δ*vc0175-176*	O1 El Tor Δ*vc0175-176*	This study
BYH256	Δ*vc0177-181*	O1 El Tor Δ*vc0177-181*	This study
BYH257	Δ*vc0182-185*	O1 El Tor Δ*vc0182-185*	This study
GS05	Δ*vc0175*	O1 El Tor Δ*vc0175 (dcdV)*	This study
WLN5105	Δ*capV*	O1 El Tor Δ*capV*	[11]
V. parahaemolyticus			
O1:Kuk str. FDA_R31	VP	Wild type	[109]
P. mirabilis			
AR379	PM	Wild type	[110]
Shigella flexneri			
PE577	Sf	Wild type	[54]
Phages			
ICP1	ICP1	Wild type	[49]
ICP3	ICP3	Wild type	[49]
Sf6	Sf6	Wild type	[111]

132

133 **Supplementary Table 2.** Plasmids Descriptions

Plasmids	Name in this Manuscript	Relevant characteristics	Source or Reference
pEVS141	pVector/pVector^{DifV}	pEVS143 without pTac; Km^r	[70]
pEVS143		Broad-host range pTac overexpression vector; Km^r	[68]
pMMB67EH	pVector^{DcdV}	Broad-host range pTac overexpression vector; Amp^r	[69]
pKAS32		Suicide vector for mutant construction, Amp^r	[67]
pET28b	pVector^{6xHis}	T7 promoter; Km^r	Novagen
pLAFR	pLAFR	pLAFR; Tet^r	[112]
pCCD7	pCCD7	pLAFR::*VSP-1*; Tet^r	[11]
pBRP353	pDncV	pMMB67EH::*dncV*; Amp^r	[11]
pCMW204	pDcdV	pEVS143::*dcdV;* Km^r	This study
pGBS87	pDcdV/pDcdV^{VC}	pMMB67EH::*dcdV*; Amp^r	This study
pGBS65	pDcdV^{6xHis}	pET28b*::dcdV-6xHis C-term*; Km^r (*only* in *E. coli* BL21(DE3))	This study
pGBS98	pDcdV^{6xHis}	pEVS143::*dcdV-6xHis C-term;* Km^r (*only* in *V. cholerae*)	This study
pGBS71	pDcdV^{E384A}	pEVS143::*dcdV-E384A;* Km^r	This study
pGBS82	pDcdV^{E384A}	pET28b::*dcdV-E384A-6xHis C-term*; Km^r (*only* for in vitro and western blot)	This study
pGBS81	pDcdV$^{C411A+C414A}$	pEVS143::*dcdV-C411A+C414A*; Km^r	This study
pGBS75	pDcdV$^{C411A+C414A}$	pET28b::*dcdV-C411A+C414A-6xHis C-term*; Km^r	This study
pGBS103	pDcdV^{S52K}	pEVS143::*dcdV-S52K;* Km^r	This study
pGBS114	pDcdV^{S52K}	pET28b::*dcdV-S52K-6xHis C-term*; Km^r	This study
pGBS106	pDcdV$^{D162A+Q163A}$	pEVS143::*dcdV-D162A+Q163A;* Km^r	This study
pGBS116	pDcdV$^{D162A+Q163A}$	pET28b::*dcdV-D162A+Q163A-6xHis C-term*; Km^r	This study
pGBS80	pIg^{222}	pEVS143::*Ig^{222}, (position in N16961 chromosome I [177,230-177,008]);* Km^r	This study
pGBS108	pIg$^{222\text{-}STOP}$	pEVS143::*ig^{222}*-1C>T, 2T>A; Km^r	This study
pGBS110	pDifV	pEVS143::*difV* (*position in N16961 chromosome I [177,181-177,008]);* Km^r	This study
pAW01	pDifV$^{RBS\text{-}less}$	pEVS143::*difV* without RBS; Km^r	This study
pGBS111	pDifV^{STOP}	pEVS143::*difV*-1A>T, 2T>A, 3G>A; Km^r	This study
pGBS118	pDifV^{InteriorSTOP}	pEVS143::*difV*-17A>T, 18T>A, 19G>A; Km^r	This study

pBYH49	pDifV$^{49\text{-}186}$	pEVS143::*difV* (49-186 NT); Km^r	This study
pBYH52	pDifV$^{49\text{-}204}$	pEVS143::*difV* (49-204 NT); Km^r	This study
pBYH53	pDifV$^{49\text{-}214}$	pEVS143::*difV* (49-214 NT); Km^r	This study
pBYH54	pDifV$^{49\text{-}218}$	pEVS143::*difV* (49-218 NT); Km^r	This study
pBYH55	pDifV$^{66\text{-}222}$	pEVS143::*difV* (66-222 NT); Km^r	This study
pBYH56	pDifV$^{86\text{-}222}$	pEVS143::*difV* (86-222 NT); Km^r	This study
pBYH57	pDifV$^{123\text{-}222}$	pEVS143::*difV* (123-222 NT); Km^r	This study
pBYH50	pNpcR_3991	pEVS143::*npcR_3991*; Km^r	This study
pGBS120	pDcdV^{E123K}	pEVS143::*dcdV*-E123K; Km^r	This study
pGBS131	pDcdV^{A126T}	pEVS143::*dcdV*-A126T; Km^r	This study
pGBS128	pDcdV^{K201R}	pEVS143::*dcdV*-K201R; Km^r	This study
pGBS129	pDcdV^{K511R}	pEVS143::*dcdV*-K511R; Km^r	This study
pGBS130	pDcdV^{Q514R}	pEVS143::*dcdV*-Q514R; Km^r	This study
pGBS124	pDcdV^{ETEC}	pEVS143::*dcdV* from *Escherichia coli* O78:H11 H10407 (ETEC); Km^r (*only* for mass spec experiment)	This study
pGBS125	pDifV^{ETEC}	pEVS143::*difV* from *Escherichia coli* O78:H11 H10407 (ETEC); Km^r	This study
pGBS126	pDcdV^{ETEC}	pMMB67EH::*dcdV* from *Escherichia coli* O78:H11 H10407 (ETEC); Amp^r	This study
pAW07	pDifV^{VP}	pEVS143::*difV* from *V. parahaemolyticus* O1:Kuk str. FDA_R31; Km^r	This study
pAW06	pDcdV^{VP}	pMMB67EH::*dcdV* from *V. parahaemolyticus* O1:Kuk str. FDA_R31; Amp^r	This study
pAW02	pDifV^{PM}	pEVS143::*difV* from *P. mirabilis* AR379; Km^r	This study
pAW04	pDcdV^{PM}	pMMB67EH::*dcdV* from *P. mirabilis* AR379	This study
pCRR01		Deletion construct for ΔVSP-1, Amp^r	This study
pCRR02		Deletion construct for ΔVSP-2, Amp^r	This study
pBYH36		Deletion construct for *Δig^{222}*, Amp^r	This study
pBYH37		Deletion construct for *Δvc0176,* Amp^r	This study
pBYH40		Deletion construct for *ΔdcdV-vc0176*, Amp^r	This study
pBYH41		Deletion construct for Δ*vc0177-vc0181,* Amp^r	This study
pBYH42		Deletion construct for Δ*vc0182-VC0185,* Amp^r	This study
pGBS88		Deletion construct for Δ*dcdV,* Amp^r	This study

134

135 **Supplementary Table 3.** Oligonucleotides used in this study.

Name	Primer use	Sequence	Reference
Vector Construction			
CMW3009	*dcdV* F[1] EcoRI + RBS[3] (pEVS143-DcdV)	GGAAACAGCCTCGACAGGCCTAGGAG GAAGCTAAATTGTTTACAATGAATAAGT CCTCCG	This study
CMW3010	*dcdV* R[2] BamHI (pEVS143-DcdV)	CATAAAGCTTGCTCAATCAATCACCGG ATCCTAGTCTTGGATGCTCTCTTC	This study
CMW3025	*dcdV* F EcoRI + RBS (pMMB67EH-DcdV)	ATTTCACACAGGAAACAGAGGAGCTAA GGAAGCTAAATTGTTTACAATGAATAAG TCCTC	This study
CMW3026	*dcdV* R BamHI (pMMB67EH-DcdV)	CCTGCAGGTCGACTCTAGAGCTAGTCT TGGATGCTCTC	This study
CMW3066	*dcdV*+6His R BamHI (pEVS143-DcdV-His_6)	GCTTGCTCAATCAATCACCGTTAGTGG TGGTGGTGGTGGTGCTCGATGTCTTGG	This study
CMW3079	Ig^{222} F EcoRI + RBS (pEVS143-Ig^{222})	CAGCCTCGACAGGCCTAGGAGGAGCT AAGGAAGCTAAACTGTTCGCAAATCAT ACTTTAG	This study
CMW3080	Ig^{222} R BamHI (pEVS143-Ig^{222}, pEVS143-DifV & pEVS143-DifV 3' end truncations and interior stop codon)	GCTTGCTCAATCAATCACCGTTACCAAT GGATTTTTTGTG	This study
CMW3081	$Ig^{222\text{-}STOP}$ F EcoRI + RBS (pEVS143- $Ig^{222\text{-}STOP}$)	CAGCCTCGACAGGCCTAGGAGGAGCT AAGGAAGCTAAATAGTTCGCAAATCAT AGTTTAG	This study
CMW3093	*dcdV* F NcoI (pET28b-DcdV-His_6)	AACTTTAAGAAGGAGATATACATGTTTA CAATGAATAAGTCCTCCGC	This study
CMW3094	*dcdV* R XhoI (pET28b-Dcdv-His_6)	CTCAGTGGTGGTGGTGGTGGTGCTCG ATGTCTTGGATGCTCTCTTCTTCACTCG ATGG	This study
CMW3102	*difV* F EcoRI + RBS (pEVS143-DifV & pEVS143-DifV 5' end truncations)	CTCGACAGGCCTAGGAGGAGCTAAGG AAGCTAAAATGATTACAAGCATTCATGA ATATAG	This study
CMW3103	*difV* F EcoRI + RBS (pEVS143-$DifV^{STOP}$)	CTCGACAGGCCTAGGAGGAGCTAAGG AAGCTAAATAAATTACAAGCATTCATGA ATATAG	This study
CMW3128	$difV^{49\text{-}186}$ F EcoRI + RBS (pEVS143-$DifV^{49\text{-}186}$)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGATTACAAG	This study
CMW2129	$difV^{49\text{-}186}$ R BamHI (pEVS143-$DifV^{49\text{-}186}$)	GCTTGCTCAATCAATCACCGGGCTCTA GCTTTCTCTTTTTTTGCGTCTTTC	This study
CMW3130	*npcR*_3991 F EcoRI + RBS (pEVS143-npcR_3991)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAACTCTCCCATAACTC	This study

CMW3131	*npcR*_3991 R BamHI (pEVS143-npcR_3991)	GCTTGCTCAATCAATCACCGTGTGCAG CACGCAAAAGATTGGCTCTAGCT	This study
CMW3162	*dcdV*^{ETEC} F EcoRI + RBS (pEVS143-DcdV^{ETEC})	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGGCTATAGCTTT GAAAAAG	This study
CMW3163	*dcdV*^{ETEC} R BamHI (pEVS143-DcdV^{ETEC})	GCTTGCTCAATCAATCACCGTTAAATCA AGTCATCTTGTTTTG	This study
CMW3164	*dcdV*^{ETEC} F EcoRI + RBS (pMMB67EH-DcdV^{ETEC})	AATTTCACACAGGAAACAGAGGAGCTA AGGAAGCTAAAATGGCTATAGCTTTGA AAAAGG	This study
CMW3165	*dcdV*^{ETEC} F BamHI (pMMB67EH-DcdV^{ETEC})	CCTGCAGGTCGACTCTAGAGTTAAATC AAGTCATCTTGTTTTGG	This study
CMW3166	*difV*^{ETEC} F EcoRI + RBS (pEVS143-DifV^{ETEC})	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGTCAAACCAATTA ACCG	This study
CMW3167	*dcdV*^{ETEC} F BamHI (pEVS143-DifV^{ETEC})	GCTTGCTCAATCAATCACCGCTAATCA AGTATTATTTCTTTCTTTAGTATTTTATC	This study
CMW3180	*difV*^{VP} F EcoRI + RBS (pEVS143-DifV^{VP})	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGGTTACAAATTTA AATG	This study
CMW3181	*difV*^{VP} R BamHI (pEVS143-DifV^{VP})	GCTTGCTCAATCAATCACCGTTACCAA CGAATTTTCTGTGCGGCTCTTAAAAG	This study
CMW3184	*dcdV*^{VP} F EcoRI + RBS (pMMB67EH-DcdV^{VP})	CAATTTCACACAGGAAACAGAGGAGCT AAGGAAGCTAAAATGGGAAAATCCTCT A	This study
CMW3185	*dcdV*^{VP} R BamHI (pMMB67EH-DcdV^{VP})	CCTGCAGGTCGACTCTAGAGTTATTCA ATAGTGGCTTCTACTTGTTGCTTTGTGA ATG	This study
CMW3189	*difV* F EcoRI (pEVS143-DifV)	ACAGCCTCGACAGGCCTAGGATGATTA CAAGCATTCATGAATATAGAAACGCTTC	This study
CMW3192	*difV*^{PM} F EcoRI + RBS (pEVS143-DifV^{PM})	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGAACGTTCAAC	This study
CMW3193	*difV*^{PM} R BamHI (pEVS143-DifV^{PM})	GCTTGCTCAATCAATCACCGTTACCAAT CTAACGTGTCTGCTACAGCTGC	This study
CMW3196	*dcdV*^{VP} F EcoRI + RBS (pMMB67EH-DcdV^{PM})	CAATTTCACACAGGAAACAGAGGAGCT AAGGAAGCTAAAATGGGTAATCC	This study
CMW3197	*dcdV*^{VP} R BamHI (pMMB67EH-DcdV^{PM})	CCTGCAGGTCGACTCTAGAGTTAACTT CTCTCTTCACCTAAACGAAGATTTAC	This study
CMW3200	*difV*$^{49-204}$ R BamHI (pEVS143-DifV$^{49-204}$)	GCTTGCTCAATCAATCACCGTGCAGCA CGCAAAAGATTG	This study
CMW3201	*difV*$^{49-214}$ R BamHI (pEVS143-DifV$^{49-214}$)	GCTTGCTCAATCAATCACCGGGATTTTT TGTGCAGCAC	This study
CMW3202	*difV*$^{49-218}$ R BamHI (pEVS143-DifV$^{49-218}$)	GCTTGCTCAATCAATCACCGCAATGGA TTTTTTGTGCAGCACGCAAAAGA	This study
CMW3203	*difV*$^{66-222}$ F EcoRI + RBS (pEVS143-DifV$^{66-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAGAATATAGAAACG	This study

CMW3204	*difV*$^{86-222}$F EcoRI + RBS (pEVS143-DifV$^{86-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATAGCGACAAAAAC	This study
CMW3205	*difV*$^{123-222}$ F EcoRI + RBS (pEVS143-DifV$^{123-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAAGACACTAGCG	This study
Site-directed Mutagenesis			
CMW3011	*dcdV* (E384A) F (pEVS143-DcdV^{E384A} & pET28b-DcdV^{E384A})	CAAGAGCGGTTCATGCTGCAATGGATT CTCTTATAGC	This study
CMW3012	*dcdV* (E384A) R (pEVS143-DcdV^{E384A} & pET28b-DcdV^{E384A})	GCTATAAGAGAATCCATTGCAGCATGA ACCGCTCTTG	This study
CMW3013	*dcdV* (C411A + C414A) F (pEVS143-DcdV$^{C411A+C414A}$)	TATATGTTACGACATATCCGGCTCACAA CGCTGCGCGACACATCGTTGCTG	This study
CMW3014	*dcdV* (C411A + C414A) R (pEVS143-DcdV$^{C411A+C414A}$)	CAGCAACGATGTGTCGCGCAGCGTTGT GAGCCGGATATGTCGTAACATATA	This study
CMW3021	*dcdV* (K55A) F (pEVS143-DcdV^{K55A})	GCTATTGGCTCTGGTGTAGCGGCATTA AAAGAGAGTTTAGTTAGTTCTCTTGAGA CATAT	This study
CMW3022	*dcdV* (K55A) R (pEVS143-DcdV^{K55A})	ATATGTCTCAAGAGAACTAACTAAACTC TCTTTTAATGCCGCTACACCAGAGCCA ATAGC	This study
CMW3104	*dcdV* (D162A + Q163A) F (pEVS143-DcdV$^{D162A+Q163A}$)	CGCATACATCATCGCGGCGTTAAAGCA CCCTGATGAAATCAAATTCC	This study
CMW3105	*dcdV* (D162A + Q163A) R (pEVS143-DcdV$^{Q162A+Q163A}$)	GGAATTTGATTTCATCAGGGTGCTTTAA CGCCGCGATGATGTATGCG	This study
CMW3110	*dcdV* (S52K) F (pEVS143-DcdV^{S52K})	CCTCTGTGGGGCTATTGGCAAAGGTGT AAAGGCATTAAAAGAGAG	This study
CMW3111	*dcdV* (S52K) R (pEVS143-DcdV^{S52K})	CTCTCTTTTAATGCCTTTACACCTTTGC CAATAGCCCCACAGAGG	This study
CMW3112	*dcdV* (S52P) F (pEVS143-DcdV^{S52P})	CCTCTGTGGGGCTATTGGCCCGGGTG TAAAGGCATTAAAAGAGAG	This study
CMW3113	*dcdV* (S52P) R (pEVS143-DcdV^{S52P})	CTCTCTTTTAATGCCTTTACACCCGGG CCAATAGCCCCACAGAGG	This study
CMW3114	*dcdV* (S52W) F (pEVS143-DcdV^{S52W})	CCTCTGTGGGGCTATTGGCTGGGGTG TAAAGGCATTAAAAGAGAG	This study
CMW3115	*dcdV* (S52K) R (pEVS143-DcdV^{S52W})	CTCTCTTTTAATGCCTTTACACCCCAGC CAATAGCCCCACAGAGG	This study
CMW3118	*difV* (interior alternative frame stop) F (pEVS143-DifV17A>T, 18T>A, 19G>A)	AAGGAAGCTAAAATGATTACAAGCATT CTAAAATATAGAAACGCTTCTAATAGCG	This study
CMW3119	*difV* (interior alternative frame stop)	CGCTATTAGAAGCGTTTCTATATTTTAG AATGCTTGTAATCATTTTAGCTTCCTT	This study

	R (pEVS143-DifV17A>T, 18T>A, 19G>A)		
CMW3124	*dcdV* (E123K) F (pEVS143-DcdV^{E123K})	GCAGCCTGTGCTATCAAAGAAATTGCG CTGG	This study
CMW3125	*dcdV* (E123K) R (pEVS143-DcdV^{E123K})	CCAGCGCAATTTCTTTGATAGCACAGG CTGC	This study
CMW3172	*dcdV* (A126T) F (pEVS143-DcdV^{A126T})	GCTATCGAAGAAATTACGCTGGAAAGA ACATTAATCTGTC	This study
CMW3173	*dcdV* (A126T) R (pEVS143-DcdV^{A126T})	GACAGATTAATGTTCTTTCCAGCGTAAT TTCTTCGATAGC	This study
Gene Deletion			
CMW2794	ΔVSP-2 up[4] F; CR02 & CR03	GTGGAATTCCCGGGAGAGCTCGGCTT GTTCACTATCGTAATAATGC	This study
CMW2795	ΔVSP-2 up R; CR02 & CR03	GGAGGGGCCACCACTGGGAGGGCACC AGATTC	This study
CMW2796	ΔVSP-2 down[5] F; CR02 & CR03	GCCCTCCCAGTGGTGGCCCCTCCCAG GT	This study
CMW2797	ΔVSP-2 down R; CR02 & CR03	AGCTATAGTTCTAGAGGTACGGGCATT AAGGTGGTGGAAACCG	This study
CMW2814	ΔVSP-1 up F; CR01 & CR03	GTGGAATTCCCGGGAGAGCTGGCTTTA CTGTTATTCGC	This study
CMW2815	ΔVSP-1 up R; CR01 & CR03	TACCATGTAGTAGCGGTATCGAGATTC C	This study
CMW2816	ΔVSP-1 down F; CR01 & CR03	GATACCGCTACTACATGGTAACGAACT CTTC	This study
CMW2817	ΔVSP-1 down R; CR01 & CR03	AGCTATAGTTCTAGAGGTACCGCTAAG TTTGTGGATGC	This study
CMW2970	Δ*vc0176* up F; BYH207	ATAACAATTTGTGGAATTCCCGGGAGA GCTGGGAATCGAATATTGAGAG	This study
CMW2971	Δ*vc0176* up R; BYH207	ATATAGTGTCTCTATTTATGGCTCATAA TCTTGAAG	This study
CMW2972	Δ*vc0176* down F; BYH207	GATTATGAGCCATAAATAGAGACACTAT ATTTAGTGTTTAATTAAC	This study
CMW2973	Δ*vc0176* down R; BYH207	TGCGCATGCTAGCTATAGTTCTAGAGG TACTATGAAACTTATTTCTATACTCTCA G	This study
CMW3035	Δ*vc0176-vc0175* up F; BYH255	ATAACAATTTGTGGAATTCCCGGGAGA GCTGGGAATCGAATATTGAGAG	This study
CMW3036	Δ*vc0176-vc0175* up R; BYH255	TTTTCCAGACTAAAGTTATGGCTCATAA TCTTGAAG	This study
CMW3037	Δ*vc0176-vc0175* down F; BYH255	GATTATGAGCCATAACTTTAGTCTGGAA AATTCACTTTTC	This study
CMW3038	Δ*vc0176-vc0175* down R; BYH255	TGCGCATGCTAGCTATAGTTCTAGAGG TACACATGGAGCATGATCAGG	This study
CMW3039	Δ*vc0177-vc0181* up F; BYH256	ATAACAATTTGTGGAATTCCCGGGAGA GCTGTTTGTATGTTTGGGGTG	This study
CMW3040	Δ*vc0177-vc0181* up R; BYH256	AATGAATTAGTATACGTATTTCTAATAC CACTAAAAACTAAG	This study

CMW3041	Δ*vc0177-vc0181* down F; BYH256	TGGTATTAGAAATACGTATACTAATTCA TTCACTGTACTTC	This study
CMW3042	Δ*vc0177-vc0181* down R; BYH256	TGCGCATGCTAGCTATAGTTCTAGAGG TACAAAGTTCTCCACAAATTTCAG	This study
CMW3043	Δ*vc0182-vc0185* up F; BYH257	ATAACAATTTGTGGAATTCCCGGGAGA GCTGCTGACTCCGGTGGCCGT	This study
CMW3044	Δ*vc0182-vc0185* up R; BYH257	CTTAGGTATACTAATTGTATTTGATATA CATAGAGGCTAGTATGGTTTCCAGAGT TTAC	This study
CMW3045	Δ*vc0182-vc0185* down F; BYH257	TGTATATCAAATACAATTAGTATACCTA AGATTCGATTTTC	This study
CMW3046	Δ*vc0182-vc0185* down R; BYH257	TGCGCATGCTAGCTATAGTTCTAGAGG TACTTCTCAGGATGTAATATTTGTG	This study
CMW3067	Δ*vc0175* up F; GS05	GTGGAATTCCCGGGAGAGCTACTATAT TTAGTGTTTAATTAACAAAAAAC	This study
CMW3068	Δ*vc0175* up R; GS05	CAGACTAAAGCCTGAAATTATGAAACTT ATTTCTATAC	This study
CMW3069	Δ*vc0175* down F; GS05	TAATTTCAGGCTTTAGTCTGGAAAATTC ACTTTTC	This study
CMW3070	Δ*vc0175* down R; GS05	AGCTATAGTTCTAGAGGTACACATGGA GCATGATCAGG	This study
CMW3071	ΔIg^{222} up F; BYH206	ATAACAATTTGTGGAATTCCCGGGAGA GCTTCTCAAAGAAGCACGTAAAAAAG	This study
CMW3072	ΔIg^{222} up R; BYH206	CAAGAATTAACGTGGTAAAGTGCGCAC ATTCTAC	This study
CMW3073	ΔIg^{222} down F; BYH206	AATGTGCGCACTTTACCACGTTAATTCT TGATTAGC	This study
CMW3074	ΔIg^{222} down R; BYH206	TGCGCATGCTAGCTATAGTTCTAGAGG TACTCATTTTCTTCTGAGGTTTC	This study
qRT-PCR			
CMW2926	*gyrA* F	TGGCCAGCCAGAGATCAAG	This study
CMW2927	*gyrA* R	ACCCGCAGCGGTACGA	This study
CMW3206	*dcdV* F	TCGACCAGTTAAAGCACCCT	This study
CMW3207	*dcdV* R	CCTTCTGTACGGATCAAGCCA	This study
CMW3208	*difV* F	GTGAATGGATATTTCGGTGGA	This study
CMW3209	*difV* R	TTGTCGCTATTAGAAGCGTT	This study
CMW3288	*ori* F	CAGGTGAACCAGCAAAATCGA	[101]
CMW3289	*ori* R	TGGTATTGAAGCTCAATGCGG	[101]
CMW3290	*ter* F	TTCAAGCTGAGGCGGATTTG	[101]
CMW3291	*ter* R	GCTCATTGGCTTCTTGTGCTT	[101]

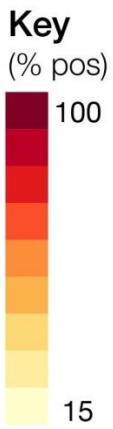
[1]F = Forward
[2]R= Reverse
[3]RBS= Ribosomal Binding Site
[4]Up= Amplifies Upstream Fragment
[5]Down= Amplifies Downstream Fragment

142 **Supplementary Table 4.** This table sorts the indicated lineages by the DcdV homolog in that
143 group with the maximum amino acid similarity to *V. cholerae* DcdV.

DcdV homologs summary table

Lineages and percentage similarities of DcdV homologs containing both DCD and PLK domains

DomArch.Gene3D	Lineage	Max%Positive
PLK+DCD	Bacteria>Proteobacteria	100.00
PLK+DCD	Bacteria>Bacteroidetes	58.80
PLK+DCD	Bacteria>Balneolaeota	56.02
PLK+DCD	Bacteria>Actinobacteria	55.63
PLK+DCD	Archaea>Thaumarchaeota	53.65
PLK+DCD	Bacteria>Firmicutes	52.27
PLK+DCD	Bacteria>Planctomycetes	52.27
PLK+DCD	Bacteria	51.88
NABP+PLK+DCD	Bacteria>Proteobacteria	51.09
PLK+DCD	Bacteria>Acidobacteria	49.70
PLK+DCD	Bacteria>Verrucomicrobia	48.69
PLK+DCD	Bacteria>Chlamydiae	45.25
PLK+DCD+NABP	Bacteria>Proteobacteria	42.48
PLK+DCD+NABP+NABP	Bacteria>Proteobacteria	39.85
PLK+PLK+DCD	Bacteria>Proteobacteria	36.75
PLK+DCD	Bacteria>Cyanobacteria	34.63
PLK+DCD	Eukaryota>Ascomycota	27.63
PLK+DCD	Eukaryota>Ciliophora	27.57
PLK+DCD	Eukaryota>Basidiomycota	25.10
PLK+DCD	Eukaryota>Chytridiomycota	23.77
PLK+DCD	Eukaryota>Mucoromycota	22.63
PLK+DCD	Eukaryota>Apicomplexa	19.96
PLK+DCD	Eukaryota>Streptophyta	19.39
PLK+Znf_CCHC+DCD	Eukaryota>Ascomycota	16.92



Abbreviations. PLK, P-loop containing nucleotide triphosphate hydrolases; DCD, Cytidine Deaminase domain 2; NABP, Nucleic acid-binding proteins; Znf_CCHC, Zinc finger CCHC-type

144

Partial Correlation Value w_{ij} of VSP-1 Genes *i* to *j* (Supplemental File 1)

	VC0175	VC0176	VC0177	VC0178	VC0179	VC0180	VC0181	VC0182	VC0183	VC0184	VC0185
VC0175	1	-0.301	-0.034	-0.459	0.147	-0.085	-0.072	0.089	-0.068	0.02	-0.099
VC0176	-0.301	1	0.145	-0.394	0.116	-0.048	-0.036	0.077	-0.055	0.028	-0.031
VC0177	-0.034	0.145	1	-0.043	0.026	-0.038	0.03	0.095	0.069	0.151	0.021
VC0178	-0.459	-0.394	-0.043	1	0.225	-0.098	-0.057	0.086	-0.11	0.027	-0.146
VC0179	0.147	0.116	0.026	0.225	1	0.501	0.303	-0.035	0.059	-0.002	0.062
VC0180	-0.085	-0.048	-0.038	-0.098	0.501	1	0.293	0.025	-0.024	0.001	-0.008
VC0181	-0.072	-0.036	0.03	-0.057	0.303	0.293	1	0.042	-0.005	0.007	-0.035
VC0182	0.089	0.077	0.095	0.086	-0.035	0.025	0.042	1	0.088	0.568	0.088
VC0183	-0.068	-0.055	0.069	-0.11	0.059	-0.024	-0.005	0.088	1	0.166	0.459
VC0184	0.02	0.028	0.151	0.027	-0.002	0.001	0.007	0.568	0.166	1	0.141
VC0185	-0.099	-0.031	0.021	-0.146	0.062	-0.008	-0.035	0.088	0.459	0.141	1

w_{ij} = -1 genes *i* and *j* never occur in the same species

w_{ij} = 0 expected co-occurrence between unrelated genes *i* and *j* drawn from a normal distribution

w_{ij} > 0.045 suggests shared biological function (Kim and Peterson 2011)

w_{ij} = 1 genes *i* and *j* always occur in the same species

Partial Correlation Value w_{ij} of VSP-2 Genes *i* to *j* (Supplemental File 2)

	VC0490	VC0491	VC0492	VC0493	VC0494	VC0495	VC0496	VC0497	VC0498	VC0502	VC0503	VC0504
VC0490	1	0.035	0.381	0.021	-0.067	0.071	0.032	-0.016	-0.124	-0.05	-0.033	0.011
VC0491	0.035	1	0.426	-0.021	-0.001	-0.008	0.014	0.008	0.014	-0.019	0.004	0
VC0492	0.381	0.426	1	0.027	0.006	0.003	0.013	0.004	0.004	0.01	-0.013	-0.021
VC0493	0.021	-0.021	0.027	1	0.008	-0.007	0.062	0.006	0.02	0.003	0.018	0.195
VC0494	-0.067	-0.001	0.006	0.008	1	0.656	0.084	0.008	-0.11	-0.022	0.03	-0.028
VC0495	0.071	-0.008	0.003	-0.007	0.656	1	0.104	0.118	0.085	-0.007	-0.009	0.02
VC0496	0.032	0.014	0.013	0.062	0.084	0.104	1	0.009	0.031	-0.039	-0.008	0.102
VC0497	-0.016	0.008	0.004	0.006	0.008	0.118	0.009	1	-0.058	-0.043	0.028	-0.004
VC0498	-0.124	0.014	0.004	0.02	-0.11	0.085	0.031	-0.058	1	-0.014	-0.028	0.028
VC0502	-0.05	-0.019	0.01	0.003	-0.022	-0.007	-0.039	-0.043	-0.014	1	0.088	0.013
VC0503	-0.033	0.004	-0.013	0.018	0.03	-0.009	-0.008	0.028	-0.028	0.088	1	-0.018
VC0504	0.011	0	-0.021	0.195	-0.028	0.02	0.102	-0.004	0.028	0.013	-0.018	1
VC0505	-0.011	0.003	0.013	0.097	-0.02	0.017	0.108	0.018	-0.022	-0.004	-0.007	0.389
VC0506	-0.006	0.001	0.002	-0.004	-0.026	0.024	0.045	0.018	-0.005	0.01	0.064	0.09
VC0507	-0.018	0.009	0.035	0.179	-0.003	0.009	0.03	-0.011	0.005	-0.01	-0.003	0.369
VC0508	0.053	0.007	-0.002	-0.029	0.028	0.111	0.073	0.06	0.017	0.123	0.098	0.003
VC0509	0.012	-0.001	-0.023	0.037	0.057	-0.02	0.216	0.023	0.058	0.095	-0.022	0.161
VC0510	0.015	0.011	0.001	-0.012	0.049	0.067	-0.005	0.085	0.064	0.128	0.174	-0.004
VC0512	-0.015	-0.006	0	0.004	-0.006	0.028	0.041	0.037	0.012	0.154	0.1	-0.002
VC0513	0.001	0.01	0.004	0.03	-0.013	-0.007	0.049	-0.008	-0.01	0.09	0.045	-0.01
VC0514	0.015	-0.007	-0.009	0.001	-0.03	0.026	-0.034	0.005	-0.052	0.012	0.1	0.028
VC0515	-0.106	0.006	0.008	0.007	-0.079	0.029	0.035	-0.065	-0.145	0.14	-0.011	0.03
VC0516	-0.018	0	0.006	0.005	0.061	0.066	0.02	0.14	-0.06	-0.046	0.204	0.021

w_{ij} = -1 genes *i* and *j* never occur in the same species

w_{ij} = 0 expected co-occurrence between unrelated genes *i* and *j* drawn from a normal distribution

w_{ij} > 0.045 suggests shared biological function (Kim and Peterson 2011)

w_{ij} = 1 genes *i* and *j* always occur in the same species

Partial Correlation Value w_{ij} of VSP-2 Genes *i* to *j* (Supplemental File 2) (con't)

	VC0505	VC0506	VC0507	VC0508	VC0509	VC0510	VC0512	VC0513	VC0514	VC0515	VC0516
VC0490	-0.011	-0.006	-0.018	0.053	0.012	0.015	-0.015	0.001	0.015	-0.106	-0.018
VC0491	0.003	0.001	0.009	0.007	-0.001	0.011	-0.006	0.01	-0.007	0.006	0
VC0492	0.013	0.002	0.035	-0.002	-0.023	0.001	0	0.004	-0.009	0.008	0.006
VC0493	0.097	-0.004	0.179	-0.029	0.037	-0.012	0.004	0.03	0.001	0.007	0.005
VC0494	-0.02	-0.026	-0.003	0.028	0.057	0.049	-0.006	-0.013	-0.03	-0.079	0.061
VC0495	0.017	0.024	0.009	0.111	-0.02	0.067	0.028	-0.007	0.026	0.029	0.066
VC0496	0.108	0.045	0.03	0.073	0.216	-0.005	0.041	0.049	-0.034	0.035	0.02
VC0497	0.018	0.018	-0.011	0.06	0.023	0.085	0.037	-0.008	0.005	-0.065	0.14
VC0498	-0.022	-0.005	0.005	0.017	0.058	0.064	0.012	-0.01	-0.052	-0.145	-0.06
VC0502	-0.004	0.01	-0.01	0.123	0.095	0.128	0.154	0.09	0.012	0.14	-0.046
VC0503	-0.007	0.064	-0.003	0.098	-0.022	0.174	0.1	0.045	0.1	-0.011	0.204
VC0504	0.389	0.09	0.369	0.003	0.161	-0.004	-0.002	-0.01	0.028	0.03	0.021
VC0505	1	0.026	0.162	0.005	0.081	-0.003	0.025	0.031	-0.022	-0.025	-0.036
VC0506	0.026	1	0.024	0.008	0.213	0.008	-0.048	0.035	-0.008	-0.009	0.044
VC0507	0.162	0.024	1	0.016	-0.023	0.003	-0.025	0.018	0.032	0.004	-0.008
VC0508	0.005	0.008	0.016	1	0.237	0.047	0.048	0.044	-0.069	0.061	0.111
VC0509	0.081	0.213	-0.023	0.237	1	0.002	-0.086	0.095	0.113	0.042	0.018
VC0510	-0.003	0.008	0.003	0.047	0.002	1	0.153	0.005	0.026	0.081	0.014
VC0512	0.025	-0.048	-0.025	0.048	-0.086	0.153	1	-0.006	0.509	0.112	0.032
VC0513	0.031	0.035	0.018	0.044	0.095	0.005	-0.006	1	0.044	0.034	-0.004
VC0514	-0.022	-0.008	0.032	-0.069	0.113	0.026	0.509	0.044	1	-0.105	0.056
VC0515	-0.025	-0.009	0.004	0.061	0.042	0.081	0.112	0.034	-0.105	1	-0.059
VC0516	-0.036	0.044	-0.008	0.111	0.018	0.081	0.014	0.032	-0.004	-0.004	1

w_{ij} = -1 genes *i* and *j* never occur in the same species

w_{ij} = 0 expected co-occurrence between unrelated genes *i* and *j* drawn from a normal distribution

w_{ij} > 0.045 suggests shared biological function (Kim and Peterson 2011)

w_{ij} = 1 genes *i* and *j* always occur in the same species